

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 09:34:33 ; Search time 22 Seconds  
(without alignments)  
436.474 Million cell updates/sec

Title: US-10-018-290A-1  
Perfect score: 904  
Sequence: 1 MKVTLMAILACLLVNSA.....INAAWKGGSKLPEMANRKK 186

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177.5	19.6	181	4	US-09-252-991A-21217
2	161	17.8	183	4	US-09-543-681A-6409
3	134.5	14.9	197	3	US-08-996-408-2
4	134.5	14.9	197	3	US-09-310-847-2
5	134.5	14.9	197	3	US-09-310-845-2
6	134.5	14.9	197	4	US-09-548-023-2
7	133.5	14.8	186	4	US-09-489-039A-12407
8	122	13.5	443	2	US-08-795-475-6
9	118.5	13.1	382	4	US-09-800-729-206
10	118	13.1	158	4	US-09-198-452A-315
11	110	12.2	8991	4	US-08-714-741-32
12	108.5	12.0	955	1	US-08-006-676B-1
13	108.5	12.0	955	2	US-08-282-845-2
14	108.5	12.0	955	2	US-08-428-414A-3
15	108.5	12.0	955	5	PCT-US94-00324-1
16	107	11.8	344	6	5210183-2
17	107	11.8	683	6	5210183-3
18	102.5	11.3	708	3	US-08-235-836C-76
19	101.5	11.2	700	1	US-07-720-589-2
20	101.5	11.2	700	2	US-08-785-130-2
21	101.5	11.2	700	3	US-08-235-836C-66
22	101.5	11.2	700	5	PCT-US92-05539-2
23	101.5	11.2	1972	4	US-08-875-435B-4
24	101	11.2	1939	4	US-09-310-187A-1
25	100.5	11.1	187	4	US-09-328-352-7281
26	100.5	11.1	391	4	US-09-800-729-208
27	100.5	11.1	468	4	US-09-328-352-6321

28	99.5	11.0	619	1	US-08-465-746-2	Sequence 2, Appli
29	99.5	11.0	619	1	US-08-214-164-2	Sequence 2, Appli
30	99.5	11.0	619	2	US-08-467-852A-3	Sequence 3, Appli
31	99.5	11.0	619	2	US-08-246-636-2	Sequence 2, Appli
32	99.5	11.0	619	2	US-08-247-491A-3	Sequence 3, Appli
33	99.5	11.0	619	2	US-08-319-795-2	Sequence 2, Appli
34	99.5	11.0	619	2	US-08-468-985-2	Sequence 2, Appli
35	99.5	11.0	619	3	US-08-312-949-2	Sequence 2, Appli
36	99.5	11.0	648	1	US-08-072-070-2	Sequence 2, Appli
37	99.5	11.0	648	1	US-08-469-434-2	Sequence 2, Appli
38	99.5	11.0	648	1	US-08-214-222-2	Sequence 2, Appli
39	99.5	11.0	648	2	US-08-467-852A-2	Sequence 2, Appli
40	99.5	11.0	648	2	US-08-468-718-2	Sequence 2, Appli
41	99.5	11.0	648	2	US-08-247-491A-2	Sequence 2, Appli
42	99.5	11.0	648	3	US-08-446-201-3	Sequence 3, Appli
43	99.5	11.0	695	1	US-08-127-499A-23	Sequence 23, Appli
44	99.5	11.0	695	1	US-08-482-847-23	Sequence 23, Appli
45	98.5	10.9	1151	4	US-09-328-352-4744	Sequence 4744, Ap

ALIGNMENTS

RESULT 1  
US-09-252-991A-21217  
; Sequence 21217, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21217  
; LENGTH: 181  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21217

Query Match	19.6%	Score 177.5;	DB 4;	Length 181;
Best Local Similarity	30.5%;	Pred. No. 9.4e-10;		
Matches	51;	Conservative 37;	Mismatches 70;	Indels 9; Gaps 4;
Qy	5	TLMAILACLLVNSAFSADFPFIGVNSQSATAMESEAKKAAQKLGSEFG--NEKTGLE 61		
Db	18	TQFVLIITAAIWAAPSAP-AEMKIAVLNYQVALLSDRAKQYAVDAERKFGQLNKNLLE 76		
Qy	62	KQAKDLQTKADDDLOAKSAAMNQAREDKQREFFLELRNFEEKSRDFAIRVQAEHTLRQY 121		
Db	77	RDKALQ---DKLVSGSKSKSQGDKREKALDFKQKARDFOFQSKELNESKAAARDMLKK 133		
Qy	122	LAEQIYLAETIAKKKGLKLVDSASGVVLEKNLITKEILRAIN 168		
Db	134	LKPKLDOAVEETIKKGGYDMVIE--RGAVVDVKFYDITRQVIERMN 178		

RESULT 2  
US-09-543-681A-6409  
; Sequence 6409, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA  
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 6409  
; LENGTH: 183  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-6409

Query Match 17.8%; Score 161; DB 4; Length 183;  
Best Local Similarity 26.2%; Pred. No. 3.8e-08;  
Matches 44; Conservative 40; Mismatches 74; Indels 10; Gaps 4;  
QY 4 KTLMAILACILVANSATFADPPIGVNSQSIAMSESAKAAKQLQSEFGNEKTQLEKQ 63  
DB 21 KLLCAAALGMAITMAGAAQADKGVVNIQVQLQVPRDAVEKQLENEFKSRDTQLNL 80  
QY 64 AKDLQTKADDLQAKSAAMNQAREDKQREFLELRFNFEKS----RDFAIRVEQAENTLR 119  
DB 81 GKALQTAVEKYQKDAPTWNTQRTANTEXDLVAKREAYAKQAQAFQDFSR--QAE--R 136  
QY 120 QYLAECIVLAETIAKKGLKLVDSAGSVWYLEKNLDTIKHLEAI 167  
DB 137 NKIMKRVLDVAVAKKEGYDVLDD--ANTVFYFADGKDITAQVQKQV 182

RESULT 3  
US-08-996-408-2  
; Sequence 2, Application US/08996408  
; Patent No. 6245338  
; GENERAL INFORMATION:  
; APPLICANT: Kyd, Jennelle  
; APPLICANT: Cripps, Allan  
; APPLICANT: Smith, Christopher  
; TITLE OF INVENTION: PEPTIDE LIBRARIES AS A SOURCE OF SYNGENES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/996,408  
; FILING DATE: 22-DEC-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01549  
; FILING DATE: 27-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9513074.6  
; FILING DATE: 27-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7116-068  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 197 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-996-408-2

Query Match 14.9%; Score 134.5; DB 3; Length 197;  
Best Local Similarity 26.8%; Pred. No. 1.6e-05;  
Matches 56; Conservative 34; Mismatches 78; Indels 41; Gaps 7;  
QY 2 KVKTLMAILACILVANSATFADPPIGVNSQSIAMSESAKAAKQLQSEFGNEKTQLE 61  
DB 6 KVTALALGI----ALASGYASAEKIAFINAGYIFQHHDPDQAVADKLDADFVKPVAEKLA 61  
QY 62 KQAKDLQTK----ADDLQAKSAAMNQAREDKQ-----REFLELRNFEE 102  
DB 62 ASKKEVDKIAAARKKVEAKVALEKODAPRLRQADIQKROBEINKLGAEDAELOKLMQE 121  
QY 103 KSRPFAIRVEQAENTLRQ-----YLAEOIYLAETIAKKGLKLVDSAGSVWYLEKNL 157  
DB 122 --QDKKQEFQAOKEKQAEERKGLDSTQATNNLARAKGYTYVLD--ANSVVFVAVEGK 177  
QY 158 DITKEILEAINAAWKGGSKLPEMANRKK 186  
DB 178 DITEVLKSI PAS-----EKAQEKK 197

RESULT 4  
US-09-310-847-2  
; Sequence 2, Application US/09310847  
; Patent No. 6254875  
; GENERAL INFORMATION:  
; APPLICANT: Kyd, Jennelle  
; APPLICANT: Cripps, Allan  
; APPLICANT: Smith, Christopher  
; TITLE OF INVENTION: PEPTIDE LIBRARIES AS A SOURCE OF SYNGENES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/310,847  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/996,408  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9513074.6  
; FILING DATE: 27-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7116-068  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 197 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-310-847-2

Query Match 14.9%; Score 134.5; DB 3; Length 197;  
Best Local Similarity 26.8%; Pred. No. 1.6e-05;  
Matches 56; Conservative 34; Mismatches 78; Indels 41; Gaps 7;

[illegible]

RESULT 6  
US-09-548-023-2  
? Sequence 2, Application US/09548023  
? Patent No. 6313281  
? GENERAL INFORMATION:  
? APPLICANT: Kyd, Jennelle  
? Crripps, Allan  
? Smith, Christopher  
? TITLE OF INVENTION: PEPTIDE LIBRARIES AS A SOURCE OF SYNGENESIS  
? NUMBER OF SEQUENCES: 5  
? CORRESPONDENCE ADDRESS:  
?

ADDRESS: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Ver  
CURRENT APPLICATION DATA:  
FILING DATE: 12-APR-2000  
CLASSIFICATION: US/09/548,023  
PRIORITY INFORMATION: <Unknown>  
PRIORITY INFORMATION: <Unknown>

```

, PRIOR APPLICATION DATA:
,
, APPLICATION NUMBER: 08/996,408
,
, FILING DATE: <Unknown>
,
, APPLICATION NUMBER: GB 9513074.6
,
, FILING DATE: 27-JUN-1995
,
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Baldwin, Geraldine F.
,
, REGISTRATION NUMBER: 31,232
,
, REFERENCE/DOCKET NUMBER: 7116-068
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: <Unknown>
,
, TELEFAX: 212-869-8864
,
, TELEX: 66441 PENNIE
,
, INFORMATION FOR SEQ ID NO: 2:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 197 amino acids
,
, TYPE: amino acid
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: protein
,
, SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-548-023-2

```

```

Query Match      14.9%; Score 134.5; DB 4; Length 197;
Best Local Similarity 26.8%; Pred. No. 1.6e-05;
Matches 56; Conservative 34; Mismatches 78; Indels 41; Gaps 7
QY 2 KVKTLSSMAILCVLIVANSASFADPPFIVGFSQSQTAMSEAAKAAQKILQSEFGNEKTQLE 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 KVTALALGI ---ALASGYASAEKIIAFINAGYIFQHHPDQAVADKLDAEFKPVAEKLA 61
QY 62 KQAKDLQTK ---ADDILQAKSAAMSNQAREDKQ-----REFLELRNFEE 102

```

```

Db 62 ASKEVDKIAAARKKVEAKVALEKDPRLRQADIQKQBEINKLGAEDAELQKLMQE 121
Qy 103 KSRDFAIRVEQAENTLRQ-----YLAEOIYLAETIAKKGLKVLDSASGVMLEKVL 157
Db 122 --QKKVGFQAKQKQAEERGLDLSIQATNNLARAKGYTYILD--ANSVFAVEGK 177
Qy 158 DITKEILEAINAAKKGSKLPEDMANRKK 186
Db 178 DITEVLKSIKAS-----EKAQEKK 197

RESULT 7
US-09-489-039A-12407
; Sequence 12407, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 2709.2004001
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12407
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12407

```

```

Query Match 14.8%; Score 133.5; DB 4; Length 186;
Best Local Similarity 25.3%; Pred. No. 1.8e-05;
Matches 42; Conservative 38; Mismatches 79; Indels 7; Gaps 4;

Qy 3 VKTSLMAILACLLVANSFADDFIGVNSOSIAMESEAAKAAQKQSEFGNEKTKOLEK 62
Db 26 VKKLLAAGLGLAMVTSQAAD-KIALVNMNSLFQVQAQKTGVSNTLENEFKRASELQR 84
Qy 63 QAKDLQTKADDLQAKSAAMNQAREDKOREFLELRNFEKSRDPAIRVEQAENTLROYL 122
Db 85 MEGDLQSKQRLQSNKPGAE---RTKLEKQVMQARQTFPSQAQAFEDQRRARSNEERGL 141
Qy 123 ABQIYLAETIAKKGLKVLDSASGVMLEKVL-DITKEILEAI 167
Db 142 VTRIQTAVQSAKQSDILVVD--ANAVAYNSSDVKDITADVLLKQV 185

```

```

RESULT 8
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/POCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-475-6

```

```

Query Match 13.5%; Score 122; DB 2; Length 443;
Best Local Similarity 33.6%; Pred. No. 0.00075;
Matches 40; Conservative 12; Mismatches 55; Indels 12; Gaps 4;

Qy 48 KLOSEFGNEKTKOLEKQAKDLQTKADDLQAKSAAMNQAREDKOREFLELRNFEKSRDF 107
Db 48 KRAEELEKAKQALDQKDLQETKLQELQDDYDLAKESTSWDRQR----LEKELEEKKEAL 103
Qy 108 AIRVEQAENTLROYLAEOIYLAETIAKKGLKVLDSAS---GSVMYLEKNLD-ITKE 162
Db 104 ELAIDQASRDYHRATA----LEKELEEKKEALELAIDQASQDYNRANVLEKELETTIRE 158

```

```

RESULT 9
US-09-800-729-206
; Sequence 206, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: PZ044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-206

```

```

Query Match 13.1%; Score 118.5; DB 4; Length 382;
Best Local Similarity 26.8%; Pred. No. 0.0014;
Matches 34; Conservative 29; Mismatches 47; Indels 17; Gaps 4;

Qy 42 AKAAQKKLQSEFGNEKTKOLEKQAKDLQTK----ADDLQAKSAAMNQAR-----EDKQ 90
Db 227 AQDVQEKLNHQLQGLAFQMKQAEELKAKISANADELRQKLVPAENVNHGLKNTGLQ 286
Qy 91 REFLELRNFEKSRDPAIRVEQAENTLROYLAEOIYLAETIAKKGLKVLDSASGV 150
Db 287 KSLLELRSHLDQVVEFRFLKVEPYGETFNALVQV----EDLRQKLG--PLAGDVVEGHL 340
Qy 151 MYLEKNL 157
Db 341 SFLEKDL 347

```

```

RESULT 10
US-09-198-452A-315
; Sequence 315, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:

```

APPLICANT: Grifais, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
FILE OF INVENTION: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 315  
LENGTH: 158  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
FEATURE:  
NAME/KEY: SITE  
LOCATION: 1...158  
OTHER INFORMATION: Xaa=unknown or other  
US-09-198-452A-315

Query Match 13.1%; Score 118; DB 4; Length 158;  
Best Local Similarity 21.3%; Pred. No. 0.00047;  
Matches 32; Conservative 43; Mismatches 69; Indels 6; Gaps 3;

QY 27 IGVNSOSIAMESEAAAKKQKLS---EFGNEKTQLEKQAKDLOTYKADDDLOAKSAAMSN 83  
Db 10 LGVNLKRCLEESDLGKTELEAKKQFVNAKIEELTSYNKLQD-EDYMESLSD 68  
QY 84 QAREDKQREYLELRNFEKSRDFAIRVEQAENTLROYLAEQIYLAETIAKKGLKLV 143  
Db 69 SASEELRKPEDLSGEYNAYOSQYQISNKNVRIKQLEQVIAESVRSKEKLEAIL 128  
QY 144 DSAGSVMYLEKNDITKEILEINAAWK 173  
Db 129 NEE--AVLAIPAGTGTTEIAILNESFKK 156

## RESULT 11

US-08-714-741-32  
Sequence 32, Application US/08714741  
Patent No. 6500613  
GENERAL INFORMATION:  
APPLICANT: Briles, David E.  
APPLICANT: McDaniel, Larry S.  
APPLICANT: Swatio, Edwin  
APPLICANT: Yother, Janet  
APPLICANT: Crain, Marilyn J.  
APPLICANT: Hollingshead, Susan  
APPLICANT: Brooks-Walter, Alexis  
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,  
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,  
TITLE OF INVENTION: PORTIONS AND PRODUCTS  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/714,741  
FILING DATE: 16-SEP-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer Esq., William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454312-2460  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8991 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
US-08-714-741-32

Query Match 12.2%; Score 110; DB 4; Length 8991;  
Best Local Similarity 24.3%; Pred. No. 0.6;  
Matches 54; Conservative 36; Mismatches 72; Indels 60; Gaps 9;

QY 7 SMAILACLIVANSAFSAADFFIGVNSOSIAMESEAAKAKKQKLSQSEFGNEKTQLE---- 61  
Db 1170 SVAILGAGLVA-----SSPTVVRAEEAPVASQSKAEKDYDTAKRDAENAKKA 1216  
QY 62 -KQAKDLOTKADDDLOAK---SAAMSNQAREDKQREYLELRNFEKSRDF----- 107  
Db 1217 LEEAKRAEQKAYDYQRIEKAKEQASLEQQ-----EANKDYQLKKYLDGGRNLNS 1272  
QY 108 --AIRVEQAENTLROYLAEQIYLAETI-----AKKGLKLVLDSSAGSVYVL--- 153  
Db 1273 VLKKEEAEKQENQAEFNKIRREIVVNPQOLEMARKEK-EVVKATESGLVTRVEEA 1331  
QY 154 EKN-----LDITKEILEINAAWKQKSGKLPENANRKK 186  
Db 1332 EKNVTDARQKLVLCNEVVLQAXXAEELESGHKLPEPNKKK 1373

## RESULT 12

US-08-006-676B-1  
Sequence 1, Application US/08006676B  
Patent No. 5411865  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven  
TITLE OF INVENTION: Diagnosis of Leishmaniasis  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jeffrey B. Oster  
STREET: 8339 SE 57th Street  
CITY: Mercer Island  
STATE: Washington  
COUNTRY: USA  
ZIP: 98040-4906  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORD for Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/006,676B  
FILING DATE: 15-JAN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oster, Jeffrey B.  
REGISTRATION NUMBER: 32,585  
REFERENCE/DOCKET NUMBER: REED-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 232 7845  
TELEFAX: (206) 236 0205  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 955 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-006-676B-1

Query Match 12.0%; Score 108.5; DB 1; Length 955;  
Best Local Similarity 26.5%; Pred. No. 0.043;

	Matches	40; Conservative	32; Mismatches	52; Indels	27; Gaps	5;
Qy	35	IAESEAAKAKKQLQSEFGNEKTLQLEKQAKDLTKKADDLQAK-----SAAMSNQAREDK	89			
Db	757	LAQLEATAAKMSAEQDRENTRTLSEQLRDSERAAELASQLESTTAAKMSAEQDRES	816			
Qy	90	QRFFLELR-RNFEKSRDFAIRVE-----QAEN-----TLRQVLAEQIYLAETIA	134			
Db	817	TRATLEQQLRDSERAAELASQLESTTAAKMSAEQDRESTRATLEQQLRESEERAAELAS	876			
Qy	135	KKGLKVLVDLSAGSVMYLEKNLDITKEILE	165			
Db	877	Q-----LESTTAAKMSAEQDRESTRATLE	900			

```

RESULT 13
US-08-282-845-2
; Sequence 2, Application US/08282845
; Patent No. 5719263
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1
; SOFTWARE: Microsoft Word for Macintosh 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/006,676
; FILING DATE: JANUARY 15, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: 5004-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-282-845-2

```

```

Query Match      12.0%; Score 108.5; DB 1; Length 955;
Best Local Similarity 26.5%; Pred. No. 0.043;
Matches 40; Conservative 32; Mismatches 52; Indels 27; Gaps 5;

Qy 35 IAMESEAAKAAQKQLQSEFGNEKTLQLEKQAKDLQTKADDLOAK-----SAAMSNQAREDK 89
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 757 LASQLEATAAKKMSAEQDRENTRTLQEQLRDSERAAELASQLESTTAAKMSAEQDRES 816
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 90 QREFLELR-RNFEKSRDFAIRVE-----QAEN-----TLQVLAECIYLAETA 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 817 TRATLEQQLRDSERAAELASQLESTTAAKMSAEQDRESITATLEQQLRESEERAAELAS 876
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 135 KKGLKLVLDASGSVMYLEKNKLDITKEILE 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      877 Q-----LESTTAAKMSAEQDRESTRATLE 900

RESULT 14
US-08-428-414A-3
; Sequence 3, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: LEISHMANIASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,414A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Kadlecsek, Ann T.
; REGISTRATION NUMBER: 39,244
; REFERENCE/DOCKET NUMBER: 210121.407
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4300
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-428-414A-3

Query Match 12.0%; Score 108.5; DB 2; Length 955
Best Local Similarity 26.5%; Pred.No. 0.043;
Matches 40; Conservative 32; Mismatches 52; Indels

QY      35 IAMESEAKAQAQKIQSEFGNEKTQLEKQAKDLQTKADDLQAK-----SAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      757 LAGLEATAAKMSAEQDRENTATLBPQURDSERAAELASQLESTTAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      90 QRBFLELR-RNFPEKSRDFAIRV-----QAEN-----TLRQVLAEE
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      817 TRATLEQQRDSEERAAELASQLESTTAAKMSAEQDRESTRATLEQQRL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      135 KKGLKVLVDSAGSVMYLEKNLDITKEILE 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      877 Q-----LESTTAAKMSAEQDRESTRATLE 900

RESULT 15
PCT-US94-00324-1
; Sequence 1, Application PC/TUS9400324
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven
; TITLE OF INVENTION: Diagnosis of Leishmaniasis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington

```

Search completed: April 27, 2004, 09:38:11  
Job time : 23 secs

16	109	121.1	2139	14	US-10-023-219-4	Sequence 39, Appl
17	107.5	11.9	364	14	US-10-100-294A-33	Sequence 3, Appl
18	107	11.6	1160	15	US-10-369-493-6564	Sequence 6564, Ap
19	106.5	11.8	1790	15	US-10-369-493-1586	Sequence 1586, Ap
20	106.5	11.8	1992	15	US-10-369-493-6527	Sequence 6527, Ap
21	106	11.7	1240	14	US-10-032-585-7319	Sequence 7319, Ap
22	105	11.6	1043	15	US-10-310-154-449	Sequence 449, App
23	104.5	11.6	475	12	US-10-425-114-46304	Sequence 46304, A
24	104.5	11.6	593	14	US-10-128-714-8303	Sequence 8303, Ap
25	104.5	11.6	1940	10	US-09-738-630-99	Sequence 99, Appl
26	104	11.5	409	12	US-10-282-122A-67318	Sequence 67318, A
27	104	11.5	1968	15	US-10-369-493-6942	Sequence 6942, Ap
28	104	11.5	1968	15	US-10-369-493-6943	Sequence 6943, Ap
29	103.5	11.4	458	12	US-10-282-122A-45264	Sequence 45264, A
30	103	11.4	250	14	US-10-234-432-25	Sequence 25, Appl
31	103	11.4	420	12	US-10-425-114-60225	Sequence 60225, A
32	102.5	11.3	708	12	US-10-369-100-76	Sequence 76, Appl
33	102	11.3	1881	14	US-10-032-585-7646	Sequence 7646, Ap
34	101.5	11.2	700	12	US-10-369-100-66	Sequence 66, Appl
35	101.5	11.2	753	15	US-10-369-493-2333	Sequence 2333, Ap
36	101	11.2	356	10	US-09-820-843A-27	Sequence 27, Appl
37	101	11.2	356	12	US-10-282-122A-77342	Sequence 77342, A
38	101	11.2	1577	15	US-10-369-493-6524	Sequence 6924, Ap
39	101	11.2	1577	15	US-10-369-493-6925	Sequence 6925, Ap
40	101	11.2	1577	15	US-10-369-493-6926	Sequence 6926, Ap
41	101	11.2	2835	9	US-09-885-535-4	Sequence 4, Appli
42	100.5	11.1	391	9	US-09-800-729-208	Sequence 208, App
43	100.5	11.1	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	11.1	440	12	US-10-282-122A-46625	Sequence 46625, A
45	100.5	11.1	879	15	US-10-369-493-1271	Sequence 1271, Ap

26	11.5	409	12	US-10-282-132A-67318	Sequence 67318, A	
27	104	11.5	1968	15	US-10-369-493-6942	Sequence 6942, A
28	104	11.5	1968	15	US-10-369-493-6943	Sequence 6943, Ap
29	103.5	11.4	448	12	US-10-282-132A-45264	Sequence 45264, A
30	103	11.4	250	14	US-10-234-432-25	Sequence 25, Appl
31	103	11.4	420	12	US-10-425-114-60225	Sequence 60225, A
32	102.5	11.3	708	12	US-10-368-100-76	Sequence 76, Appl
33	102	11.3	1881	14	US-10-034-585-7646	Sequence 7646, Ap
34	101.5	11.2	700	12	US-10-369-100-66	Sequence 66, Appl
35	101.5	11.2	753	15	US-10-369-493-2333	Sequence 2333, Ap
36	101	11.2	356	10	US-09-820-843A-27	Sequence 27, Appl
37	101	11.2	356	12	US-10-282-132A-77342	Sequence 77342, A
38	101	11.2	1577	15	US-10-369-493-6324	Sequence 6324, Ap
39	101	11.2	1577	15	US-10-369-493-6325	Sequence 6325, Ap
40	101	11.2	1577	15	US-10-369-493-6326	Sequence 6326, Ap
41	101	11.2	2835	9	US-09-885-535-4	Sequence 4, Appl
42	100.5	11.1	391	9	US-09-800-729-208	Sequence 208, App
43	100.5	11.1	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	11.1	440	12	US-10-282-132A-46625	Sequence 46625, A
45	100.5	11.1	879	15	US-10-369-493-13271	Sequence 1271, A

30	11.4	250	14	US-10-234-432-25	Sequence 25, Appl
31	103	420	12	US-10-425-114-60225	Sequence 60225, A
32	102.5	708	12	US-10-369-100-76	Sequence 76, Appl
33	102	1381	14	US-10-032-585-7646	Sequence 7646, Ap
34	101.5	700	12	US-10-363-100-66	Sequence 66, Appl
35	101.5	753	15	US-10-369-493-2333	Sequence 2333, Ap
36	101	356	10	US-09-820-843A-27	Sequence 27, Appl
37	101	356	12	US-10-283-132A-77342	Sequence 77342, A
38	101	1577	15	US-10-369-493-6924	Sequence 6924, Ap
39	101	1577	15	US-10-369-493-6925	Sequence 6925, Ap
40	101	1577	15	US-10-369-493-6926	Sequence 6926, Ap
41	101	2835	9	US-09-885-533-4	Sequence 4, Appl
42	100.5	391	9	US-09-800-729-208	Sequence 208, App
43	100.5	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	440	12	US-10-283-132A-46625	Sequence 46625, A
45	100.5	879	15	US-10-369-493-1371	Sequence 1271, Ap

33	11.3	1881	14	US-10-032-585-7646	Sequence 7646, App	
34	101.5	11.2	700	12	US-10-369-100-66	Sequence 66, Appl
35	101.5	11.2	753	15	US-10-369-493-2333	Sequence 2333, Ap
36	101	11.2	356	10	US-09-820-843A-27	Sequence 27, Appl
37	101	11.2	356	12	US-10-282-132A-77342	Sequence 77342, A
38	101	11.2	1577	15	US-10-369-493-6324	Sequence 6324, Ap
39	101	11.2	1577	15	US-10-369-493-6325	Sequence 6325, Ap
40	101	11.2	1577	15	US-10-369-493-6326	Sequence 6326, Ap
41	101	11.2	2835	9	US-09-885-533-4	Sequence 4, Appl
42	100.5	11.1	395	9	US-09-800-729-208	Sequence 208, App
43	100.5	11.1	395	9	US-09-887-107-35	Sequence 35, Appl
44	100.5	11.1	40	12	US-10-282-132A-46625	Sequence 46625, A
45	100.5	11.1	879	15	US-10-369-493-1371	Sequence 1271, Ap

35	101.5	753	15	US-10-369-493-2333	Sequence 233, Appl
36	101.1	356	10	US-09-820-843A-27	Sequence 27, Appl
37	101.1	356	12	US-10-283-132A-77342	Sequence 77342, A
38	101.1	1577	15	US-10-369-493-6324	Sequence 6324, Ap
39	101.1	1577	15	US-10-369-493-6325	Sequence 6325, Ap
40	101.1	1577	15	US-10-369-493-6326	Sequence 6326, Ap
41	101.1	2835	9	US-09-885-535-4	Sequence 4, Appl
42	100.5	395	9	US-09-807-729-208	Sequence 208, App
43	100.5	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	140	12	US-10-283-132A-46625	Sequence 46625, A
45	100.5	879	15	US-10-369-493-1371	Sequence 1271, Ap

37	101	11.2	356	12	US-10-283-132A-77342	A Sequence 77342, A
38	101	11.2	1577	15	US-10-369-493-6924	A Sequence 6924, Ap
39	101	11.2	1577	15	US-10-369-493-6925	A Sequence 6925, Ap
40	101	11.2	1577	15	US-10-369-493-6926	A Sequence 6926, Ap
41	101	11.2	2835	9	US-09-885-535-4	A Sequence 4, Appl
42	100.5	11.1	391	9	US-09-800-729-208	A Sequence 208, App
43	100.5	11.1	395	9	US-09-987-107-35	A Sequence 35, Appl
44	100.5	11.1	440	12	US-10-283-132A-46625	A Sequence 46625, A
45	100.5	11.1	879	15	US-10-369-493-1371	A Sequence 1271, Ap

39	101	11.2	1577	15	US-10-369-493-6925	Sequence 6925, Ap
40	101	11.2	1577	15	US-10-369-493-6926	Sequence 6926, Ap
41	101	11.2	2835	9	US-09-885-535-4	Sequence 4, Appl1
42	100.5	11.1	391	9	US-09-800-729-208	Sequence 208, App
43	100.5	11.1	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	11.1	440	12	US-10-283-122A-46625	Sequence 46625, A
45	100.5	11.1	879	15	US-10-369-493-1371	Sequence 1271, Ap

41	101	2835	9	US-09-885-535-4	Sequence 4, Appli
42	100.5	391	9	US-09-800-729-208	Sequence 208, App
43	100.5	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	440	12	US-10-282-122A-46625	Sequence 46625, A
45	100.5	879	15	US-10-369-493-1371	Sequence 1271, Ap

43	100.5	11.1	395	9	US-09-987-107-35	Sequence 35, Appl
44	100.5	11.1	440	12	US-10-282-122A-46625	Sequence 46625, A
45	100.5	11.1	879	15	US-10-369-493-1271	Sequence 1271, Ap

45	100.5	11.1	879	15	US-10-369-493-1271	Sequence 1271, Ap
----	-------	------	-----	----	--------------------	-------------------



## RESULT 2

US-09-741-849-2  
; Sequence 2, Application US/09741849  
; Patent No. US20020099188A1  
; GENERAL INFORMATION:  
; APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN  
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there  
; FILE REFERENCE: 77813-39  
; CURRENT APPLICATION NUMBER: US/09/741,849  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 60/171,539  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-741-849-2

Query Match 14.5%; Score 131; DB 9; Length 171;  
Best Local Similarity 22.7%; Pred. No. 0.00081;  
Matches 37; Conservative 44; Mismatches 76; Indels 6; Gaps 3;  
QY 14 LLVANSATSDPPIGVFNQSQTAMSEAAKAAQKQLQS---EFGNEKTQLEKQAKDLQTK 70  
Db 10 LLVLGSTAAHANLGYVNLKRCLEESDLGKKEETELEAMKQOFVKNAEKIEBELTSIYNK 69  
QY 71 ADDLQAKSAAMNQAREDKQREFLELRNFEKSRDFAIRVEQAENTLROYLAEOIYLAA 130  
Db 70 LQD-EDYNESLSDSASELRKPFELSGEYNAQSYQYQSQINQSNVVKRIQKLIQEVKIAA 128  
QY 131 ETIAKKKGLKLVDSAGSVMYLEKNLDITKEILEAINAAWK 173  
Db 129 ESVRSKEKLEALINEE--AVLAIAPGTDKTTIILAILNESPK 169

## RESULT 3

US-10-312-273-77  
; Sequence 77, Application US/10312273  
; Publication No. US20040005667A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: IMMUNISATION AGAINST CHLAMYDIA PNEUMONIAE  
; FILE REFERENCE: P025035WO  
; CURRENT APPLICATION NUMBER: US/10/312,273  
; CURRENT FILING DATE: 2002-12-20  
; PRIOR APPLICATION NUMBER: 0016363.4  
; PRIOR FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 0017047.2  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 0017983.8  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: 0019368.0  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 0020440.4  
; PRIOR FILING DATE: 2000-08-18  
; PRIOR APPLICATION NUMBER: 0022583.9  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 0027549.5  
; PRIOR FILING DATE: 2000-11-10  
; PRIOR APPLICATION NUMBER: 0031706.5  
; PRIOR FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 664  
; SOFTWARE: SeqWin99, version 1.02  
; SEQ ID NO 77  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-10-312-273-77

Query Match 14.5%; Score 131; DB 15; Length 171;  
Best Local Similarity 22.7%; Pred. No. 0.00081;

Matches 37; Conservative 44; Mismatches 76; Indels 6; Gaps 3;  
QY 14 LLVANSATSDPPIGVFNQSQTAMSEAAKAAQKQLQS---EFGNEKTQLEKQAKDLQTK 70  
Db 10 LLVLGSTAAHANLGYVNLKRCLEESDLGKKEETELEAMKQOFVKNAEKIEBELTSIYNK 69  
QY 71 ADDLQAKSAAMNQAREDKQREFLELRNFEKSRDFAIRVEQAENTLROYLAEOIYLAA 130  
Db 70 LQD-EDYNESLSDSASELRKPFELSGEYNAQSYQYQSQINQSNVVKRIQKLIQEVKIAA 128  
QY 131 ETIAKKKGLKLVDSAGSVMYLEKNLDITKEILEAINAAWK 173  
Db 129 ESVRSKEKLEALINEE--AVLAIAPGTDKTTIILAILNESPK 169

## RESULT 4

US-08-325-278-6  
; Sequence 6, Application US/08325278  
; Publication No. US20030027283A1  
; GENERAL INFORMATION:  
; APPLICANT: Bivrock, Lars  
; APPLICANT: Sjvbring, Ulf  
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325,278  
; FILING DATE: 26-OCT-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMasters, David D.  
; REGISTRATION NUMBER: 33,963  
; REFERENCE/DOCKET NUMBER: 450023.401  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 443 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-325-278-6

Query Match 13.5%; Score 122; DB 8; Length 443;  
Best Local Similarity 33.6%; Pred. No. 0.017;  
Matches 40; Conservative 12; Mismatches 55; Indels 12; Gaps 4;  
QY 48 KLOSEFGNEKTQLEKQAKDLQTKADDLQAKSAAMNQAREDKQREFLELRNFEKSRDF 107  
Db 48 KRAEELEKAKQALEDKQREFLELRNFEKSRDFLELRNFEKSRDF 103  
QY 108 AIRVEQAENTLROYLAEOIYLAAETIAKKKGLKLVDSAS---GSVMYLEKNLD-ITKE 162  
Db 104 ELAQASRDYHRATA---LEKELEBKKALELAIDQASQDYNRANVLEKLETTIRE 158

## RESULT 5

US-09-800-729-206  
; Sequence 206, Application US/09800729  
; Patent No. US20020068319A1  
; GENERAL INFORMATION:



Db 283 KQEAQKTLQDETKEYAKNRAARTEVEKKIQEPREDEVEQDAEIA-EAREDLDAKKKVL 341

QY 95 ELERNFEKSRDFAIRVEQENTLROYLAQIYLAETIAKKGLKLVLSASGSMYLE 154

Db 342 E-----FEK-----IRECEQSIRKKTSEKKYMERITIVNAKNEVRILLEK-----Q 382

QY 155 KNLDTITKEILEAIN 168

Db 383 GNQDLTKRLTKVEN 396

RESULT 9

US-10-369-493-5581

; Sequence 5581, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 5581

; LENGTH: 1130

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-10-369-493-5581

Query Match 12.6%; Score 113.5; DB 15; Length 1130;

Best Local Similarity 23.9%; Pred. No. 0.31;

Matches 32; Conservative 34; Mismatches 45; Indels 23; Gaps 5;

QY 38 ESEAAKAAQKQKQSEFNGEKTQLEKQAKDLQTKADDLOAKSAAMNQARED---KOREFL 94

Db 283 KQEAQKTLQDETKEYAKNRAARTEVEKKIQEPREDEVEQDAEIA-EAREDLDAKKKVL 341

QY 95 ELERNFEKSRDFAIRVEQENTLROYLAQIYLAETIAKKGLKLVLSASGSMYLE 154

Db 342 E-----FEK-----IRECEQSIRKKTSEKKYMERITIVNAKNEVRILLEK-----Q 382

QY 155 KNLDTITKEILEAIN 168

Db 383 GNQDLTKRLTKVEN 396

RESULT 10

US-10-369-493-18546

; Sequence 18546, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 18546

; LENGTH: 1190

; TYPE: PRT

; ORGANISM: Halobacterium sp. NRC-1

US-10-369-493-18546

Query Match 12.6%; Score 113.5; DB 15; Length 1190;

Best Local Similarity 27.3%; Pred. No. 0.33;

Matches 41; Conservative 35; Mismatches 51; Indels 23; Gaps 7;

QY 34 SIAMESEAAKAAQKQKQSEFNGEKTQLEKQAKDLQTKADDLOAKSAAMNQ--AREDKQ 90

Db 842 SLTLEKEYAEDAVEDLQADVADAESKADQAR-IEALNGDIEAKQAEIAEKAIVEDLE 900

QY 91 REFLELRNFEKSRDFAIRVEQENTLROYLAQIYLAETIAKKGLK 140

Db 901 AELADLKRDEERKADLSEAEDARDEQAAAVEDARHRL-----ERLQAAQTLSSEVA-- 953

QY 141 LVLDASGSMYLE-KNLDTITKEILEAIN 169

Db 954 -ELDDAVGEYDPAEIPDADVVAENVERLGA 982

RESULT 11

US-10-221-278-273

; Sequence 273, Application US/10221278

; Publication No. US20040034208A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 21272-045

; CURRENT APPLICATION NUMBER: US/10/221,278

; CURRENT FILING DATE: 2002-09-06

; PRIOR APPLICATION NUMBER: 09/693,267

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 09/665,363

; PRIOR FILING DATE: 2000-09-19

; PRIOR APPLICATION NUMBER: 09/616,847

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 09/596,193

; PRIOR FILING DATE: 2000-06-17

; PRIOR APPLICATION NUMBER: 09/574,454

; PRIOR FILING DATE: 2000-05-19

; PRIOR APPLICATION NUMBER: 09/519,705

; PRIOR FILING DATE: 2000-03-07

; NUMBER OF SEQ ID NOS: 752

; SEQ ID NO 273

; LENGTH: 876

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-221-278-273

Query Match 12.4%; Score 112.5; DB 12; Length 876;

Best Local Similarity 29.2%; Pred. No. 0.28;

Matches 42; Conservative 24; Mismatches 55; Indels 23; Gaps 6;

QY 35 IAMESEAAK--AQKKQKQSEFNGEKTQLEKQAKDLQTKADDLOAKSAAMNQAR--EDKQ 90

Db 144 VCLEGHQVKLNAAEEMQQL--LSRTSLETKDLMTEVSELKULVGMKEKQREBEKQ 202

QY 91 REFLELRNFEKSRDFAIRVEQENTLROYLAQIYLAETIAKK 137

Db 203 RKABELL-----QELRHLLKIKVELENERNOYEWKLNKATKBAVQIQEQVALKDAEIBRLH 258

QY 138 GLKLVLSASGSMYLEKNLDTITK 161

Db 259 S-QLSRTAALHSESHTERDQEIQR 281

RESULT 12

US-10-291-172-273

; Sequence 273, Application US/10291172

; Publication No. US20030228584A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 09:33:33 ; Search time 20 Seconds  
(without alignments)  
894.581 Million cell updates/sec

Title: US-10-018-290A-1

Perfect score: 904

Sequence: 1 MKVKTLSMAILACLLVANS.....INAAWKGGKGLPEMANRKK 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	186	20.6	165	2	AD0129	cationic 19 kDa ou
2	179.5	19.9	168	2	G83190	probable outer mem
3	169	18.7	143	2	SI9728	outer membrane pro
4	163	18.0	162	2	H70314	hypothetical prote
5	137.5	15.2	169	2	A82039	outer membrane pro
6	136.5	15.1	197	2	E64102	skp protein - Haem
7	134.5	14.9	161	1	DNEC17	outer membrane pro
8	134.5	14.9	161	2	D90651	histone-like prote
9	134.5	14.9	161	2	D85502	hypothetical prote
10	134	14.8	166	2	F81228	outer membrane pro
11	133	14.7	166	2	E82000	probable outer mem
12	131.5	14.5	161	1	S09104	outer membrane pro
13	131.5	14.5	161	2	AC0530	outer membrane pro
14	131	14.5	171	2	E72094	cationic outer mem
15	131	14.5	171	2	E86528	ompH-like outer me
16	128.5	14.2	484	2	S35401	M1 protein precurs
17	124.5	13.8	227	2	S01260	M1 protein precurs
18	124.5	13.8	484	2	S46489	M1 protein precurs
19	123.5	13.7	484	2	S34978	M1.1 protein precu
20	116.5	12.9	697	2	T07111	MAR binding filame
21	114.5	12.7	630	2	T38023	probable transcrip
22	114	12.6	1298	2	T24480	hypothetical prote
23	113.5	12.6	1130	2	T20288	hypothetical prote
24	113.5	12.6	1190	2	E84193	chromosome segrega
25	112.5	12.4	1940	1	A24922	myosin heavy chain
26	112	12.4	173	2	C71539	probable (ompH-lik
27	112	12.4	279	2	D71433	hypothetical prote
28	111.5	12.3	173	2	A81694	cationic outer mem
29	111.5	12.3	389	2	A43715	M49 protein precur

30	110.5	12.2	384	2	S49550	M-like protein emm
31	109.5	12.1	365	2	B54128	Fe-binding protein
32	109	12.1	532	2	S47341	skp protein - Past
33	109	12.1	532	2	S54871	M protein - Strept
34	109	12.1	587	2	JC1419	Fc gamma (IgG) rec
35	108.5	12.0	454	2	S43556	plasminogen-bindin
36	108.5	12.0	472	2	S43554	plasminogen-bindin
37	108.5	12.0	955	2	A47334	LeKin kinesin-rela
38	107.5	11.9	594	2	D95286	hypothetical prote
39	107	11.8	386	2	S54858	M protein precurs
40	107	11.8	1164	2	T24806	hypothetical prote
41	107	11.8	1175	2	C35815	myosin heavy chain
42	107	11.8	1201	2	A35815	myosin heavy chain
43	107	11.8	1201	2	B35815	myosin heavy chain
44	107	11.8	2385	2	A32491	myosin heavy chain
45	107	11.8	2411	2	B32491	myosin heavy chain

## ALIGNMENTS

### RESULT 1

AD0129

cationic 19 kDa outer membrane protein precursor ompH [imported] - Yersinia pestis (s  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AD0129

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AD0129

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-165 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC89895.1; PID:g15979120; GSPDB:GN00175

C;Genetics:

A;Gene: ompH

C;Superfamily: DNA-binding 17K protein

Query Match	20.6%;	Score 186;	DB 2;	Length 165;
Best Local Similarity	30.7%;	Pred. No. 8.9e-06;		
Matches	51;	Conservative	33;	Mismatches 76;
				Indels 6;
				Gaps 3;
Qy	4	KTLSMAILACLLVANS	AFSADFFIGVNSOSIAMESAAXAAQKQLQSEFGNEKIQLEKQ	63
Db	3	KWLCASLGLAASASVQAADKIAIVNVSIFQQLPAREAVAKQLENEFKGRATELQGM	62	
Qy	64	AKDLQTKADDLQAKSAAMNQAREDKOREFLELRNFEEKSRDFA--IRVEQAENTLRQY	121	
Db	63	ERDLQTKMQLQRDGSTMKASDRTKLENEVVKQRETFTSKAQAFQONRRRQAE--RNK	120	
Qy	122	LAEQIYLAETIAAKKGLKVLDSAGSVMYLEKNLDTREILFAI	167	
Db	121	ILSRIQDVKSVATKGGYDVVID--ANAVAYADSSKDIADVLKQV	164	

### RESULT 2

G83190

probable outer membrane protein precursor PA3647 [imported] - Pseudomonas aeruginosa  
C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: G83190

A;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-168 <STO>

A:Cross-references: GB:AE004784; GB:AE004091; NID:G9949799; PIDN:AAG07035.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3647

Query Match 19.9%; Score 179.5; DB 2; Length 168;

Best Local Similarity 30.4%; Pred. No. 2.5e-05;

Matches 52; Conservative 38; Mismatches 72; Indels 9; Gaps 4;

QY 1 MKVKTLSMAILLVANSAFSADFPPIGVFNSQSIAESEAAKAAQKQLQSEFG---NEK 57

DB 1 MRKFTQFVLITAAIMAAPSAF-AEMKIAVLNYQMALLESDAKQYAVDAEKFKGQLNKL 59

QY 58 TQLEKQADLOTKADDLOAKSAAMNQAREDKQREFLELRNFEKSRDFAIRVQAEHT 117

DB 60 KNLERDAKALQ---DKLVNSGSKMSGGREKAELDFKQKARDQFQSKELNESKAAADR 116

QY 118 LRQYLAEQIYLAETIAKKKGLKLVDSAGSVWYLEKNLDTITKEILEAIN 168

DB 117 MLKLLKPKLDQAVETIKGGYDMVIE--RGAVDVKPYDITRQVIERMN 165

RESULT 3

S19728

outer membrane protein ompH - Yersinia pseudotuberculosis (fragment)

C:Species: Yersinia pseudotuberculosis

C:Date: 13-Jan-1995 #sequence\_revision 22-May-1998 #text\_change 17-Mar-1999

C:Accession: S19728

R:Vuorio, R.; Hirvas, L.; Raybourne, R.B.; Yu, D.T.Y.; Vaara, M.

Biochim. Biophys. Acta 1129, 124-126, 1991

A:Title: The nucleotide and deduced amino acid sequence of the cationic 19 kDa outer mem

A:Reference number: S19728; MUID:92096452; PMID:1756172

A:Accession: S19728

A:Molecule type: DNA

A:Residues: 1-143 <VUC>

A:Cross-references: EMBL:M73247

C:Genetics:

A:Gene: ompH

C:Superfamily: DNA-binding 17K protein

C:Keywords: membrane protein

Query Match 18.7%; Score 169; DB 2; Length 143;

Best Local Similarity 30.8%; Pred. No. 0.0001;

Matches 44; Conservative 30; Mismatches 63; Indels 5; Gaps 3;

QY 27 IGVFNSQSIAESEAAKAAQKQLQSEFGNEKTQLEKQAKDLQTKADDLOAKSAAMNQAR 86

DB 4 IAIENVSSIFQOLPAREAVAKQLENEFKGRATELGQMERDLQTKMQLQRDGSIMKASDR 63

QY 87 EDKQREFLELRNFEKSRDFA--LRVQAEHTLROYLAEQIYLAETIAKKKGLKLVLD 144

DB 64 TKLENVKQRETFTSKAQAFQDNRKQAE--RNKILSRIQDRAVKSVAIKGGYDVVID 121

QY 145 SAGSVWYLEKNLDTITKEILEAI 167

DB 122 --ANAVAYADGSKDITADVLRQV 142

RESULT 4

H70314

hypothetical protein aq\_157 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C:Accession: H70314

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: H70314

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-162 <AQF>

A:Cross-references: GB:AB000676; NID:G2982884; PIDN:AAC06512.1; PID:G2982893; GB:AB00

A:Experimental source: strain VFS

C:Genetics:

A:Gene: aq\_157

Query Match 18.0%; Score 163; DB 2; Length 162;

Best Local Similarity 28.3%; Pred. No. 0.0003;

Matches 45; Conservative 37; Mismatches 61; Indels 16; Gaps 4;

QY 15 LVANSFASDPPIGVFNSQSIAESEAAKAAQKQLQSEFGNEKTQLEKQAKDLQTKADDL 74

DB 14 LIAGISFLDF--ACVDINKVIRSKFTAKAQTELKXE-----LEKYQKLIQEKQKL 64

QY 75 QA-----KSAAMNQAREDKQREFLELRNFEKSRDFAIRVQAEHTLROYLAEQIYLA 129

DB 65 EALKKSLESKALSKEKAKKEIQLEDELRLKQVRAQSXLRSKKAELERKWFVKVI 124

QY 130 AETIAKKKGLKLVDSAGSVWYLEKNLDTITKEILEAIN 168

DB 125 VESTAKKKKIKAVFD--CNSMLYWDKKIDITNEVLKELD 161

RESULT 5

A82099

outer membrane protein OmpH VC2251 [imported] - Vibrio cholerae (strain N16961 serogr

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: A82099

R:Heidelberg, J.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers

l., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: A82099

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-169 <HEI>

A:Cross-references: GB:AB004297; GB:AB003852; NID:99656810; PIDN:AAF95395.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2251

A:Map position: 1

C:Superfamily: DNA-binding 17K protein

Query Match 15.2%; Score 137.5; DB 2; Length 169;

Best Local Similarity 24.0%; Pred. No. 0.015;

Matches 40; Conservative 46; Mismatches 78; Indels 3; Gaps 2;

QY 1 MKVKTLSMAILLVANSFASDPPIGVFNSQSIAESEAAKAAQKQLQSEFGNEKTQL 60

DB 5 IKAASLGLIILUSSSMANAAEAAQ-KIYINTAQVQALPQREVLQKMQEFGKDKAAEL 63

QY 61 EKOAKDLQTKADDLOAKSAAMNQAREDKQREFLELRNFEKSRDFAIRVQAEHTLRQ 120

DB 64 QAIQADAKTKIEKLKEDQLMGQDEVEKLRLEIQDLSKYKIKAAQALEQASAREBEKQ 123

QY 121 YLAEOIYLAETIAKKKGLKLVDSAGSVWYLEKNLDTITKEILEAI 167

DB 124 KLFVKVIQDAVKVKAKEGYDIVLDTSS--SMQYGRPEHNLSEKVIKAI 168

RESULT 6

E64102

skp protein - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999

C:Accession: E64102

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidmar

; D.N.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

[illegible]



4.



Db 129 ESVRSKEKLEAILNEE--AVLAIAPGTDKTTETIAILNESFKK 169

Search completed: April 27, 2004, 09:37:36  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 09:32:28 ; Search time 17 seconds  
(without alignments)  
569.709 Million cell updates/sec

Title: US-10-018-290A-1

Perfect score: 904

Sequence: 1 MKVKTLSSMAILCLLVANSA.....INAAKKGGSKLPEMANRKK 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	186	20.6	165	1	OMPH_YERPE
2	179.5	19.9	168	1	OMPH_PSEAE
3	169	18.7	143	1	OMPH_YERPS
4	163	18.0	162	1	Y157_AQUAE
5	158.5	17.5	164	1	OMPH_YEREN
6	137.5	15.2	169	1	OMPH_VIRCH
7	136.5	15.1	127	1	OM26_HABIN
8	134.5	14.9	161	1	HLPA_ECOLI
9	131.5	14.5	161	1	OMPH_SALTY
10	131	14.5	171	1	OMPH_CHLPN
11	118.5	13.1	382	1	APA4_PIG
12	116.5	12.9	697	1	MFPI_LYCES
13	112.5	12.4	1940	1	MYH3_RAT
14	112	12.4	173	1	OMPH_CHLTR
15	111.5	12.3	173	1	OMPH_CHLMU
16	111.5	12.3	389	1	M49_STRPY
17	109	12.1	193	1	OM25_PASMU
18	108.5	12.0	955	1	KINL_LEICH
19	108.5	12.0	1976	1	MYH4_RAT
20	108	11.9	1191	1	SMC2_MOUSE
21	107	11.8	1962	1	MYSA_DROME
22	107	11.8	4473	1	PLE1_CRIGR
23	106.5	11.8	283	1	TPM1_LOCM1
24	106.5	11.8	1790	1	USO1_YEAST
25	106.5	11.8	1969	1	MYSA_CABEL
26	106.5	11.8	1976	1	MYH4_BOVIN
27	105	11.6	722	1	MFPI_TOBAC
28	105	11.6	790	1	EEA1_MOUSE
29	105	11.6	1938	1	MYH5_AQOIR
30	104.5	11.6	1940	1	MYH3_HUMAN
31	104	11.5	407	1	M21_STRPY
32	104	11.5	1203	1	SMC2_XENLA
33	104	11.5	1947	1	MYSC_CABEL

34	103.5	11.4	386	1	ARPA_STRPY
35	103.5	11.4	1197	1	SMC2_HUMAN
36	103	11.4	458	1	VIMI_XENLA
37	103	11.4	816	1	HUNB_DROVI
38	103	11.4	1976	1	MYHA_HUMAN
39	102.5	11.3	485	1	RNF6_HUMAN
40	102.5	11.3	1938	1	MYHD_HUMAN
41	102	11.3	284	1	TPM2_DROME
42	102	11.3	2230	1	GOA4_HUMAN
43	101.5	11.2	161	1	TPM2_YEAST
44	101.5	11.2	1191	1	CING_MOUSE
45	101.5	11.2	1972	1	MYHB_RABIT

## ALIGNMENTS

RESULT 1  
OMPH\_YERPE  
ID\_OMPHERPE STANDARD; PRT; 165 AA.  
AC PS8607; 2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cationic 19 kDa outer membrane protein precursor.  
GN OMPH OR YPO1053 OR Y3126.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titchell R.W., Holden M.T.G.,  
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holtroyd S., Jørgensen K., Karsyshev A.V.,  
RA Leather S., Mould S., Oyston P.C.F., Quail M.A., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
RA "Genome sequence of Yersinia pestis, the causative agent of plague."  
RN Nature 413:523-527(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KIM5 / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RP "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AJ414146; CAC89895.1; -  
CC EMBL; AE013913; AA866676.1; ALT\_INIT.  
CC PIR; AD0129; AD0129.  
CC InterPro; IPR005632; OmpH.  
CC Pfam; PF03938; OmpH; 1.  
CC Outer membrane; Signal; Complete proteome.  
CC SIGNAL 1 22 POTENTIAL.  
CC CHAIN 23 165 CATIONIC 19 kDa OUTER MEMBRANE PROTEIN.  
CC FT

PI3050 streptococ  
O85347 homo sapien  
P24789 xenopus lae  
PI3361 drosophila  
P35580 homo sapien  
O76064 homo sapien  
O9ukx3 homo sapien  
P09491 drosophila  
Q13439 homo sapien  
P40414 saccharomyc  
P59242 mus musculu  
P35748 oryctolagus

```
SQ SEQUENCE 165 AA; 18279 MW; 982D3BEF3A66FB6C CRC64;
Query Match 20.6%; Score 186; DB 1; Length 165;
Best Local Similarity 30.7%; Pred. No. 9.7e-06;
Matches 51; Conservative 33; Mismatches 76; Indels 6; Gaps 3;

QY 4 KTLMAILACLIVANSAPFPIGVFNSQSITAMSEAAKAAQKLOSEFGNEKTQLEBKQ 63
Db 3 KWLCAASLGLAASASVQADKIAIVNVSSIFQQLPAREAVAKQLENEFKGRATELQCM 62

QY 64 AKDLQTKADDDLOAKSAANSQAREDKQREFLELRNFEEKSRDFA--IRVEQAENTLRQY 121
Db 63 ERDLQTKQKLQDQSGTMTKASDRTKLENEVMKQRETFSTKAQAFQDNRRQAEE--RNK 120

QY 122 LAEQIYLAETIAKKGLKVLDSAGSGVMYLEKNLDTITKEILEAI 167
Db 121 ILSRIQDAVSKVATGGYDVVID--ANAVAYADSSKDTADVLKQV 164

RESULT 2
OMPH_PSEAE
ID _OMPH_PSEAE STANDARD; PRT; 168 AA.
AC Q9HX5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE OMPH-like protein precursor.
GN PA3647.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004784; AAG07035.1; --
DR PIR; G83190; G83190.
DR InterPro; IPR005632; OmpH.
DR Pfam; PF03938; OmpH; 1.
KW Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 168 OMPH-LIKE PROTEIN.
FT CHAIN 23 168 OMPH-LIKE PROTEIN.
SQ SEQUENCE 168 AA; 19090 MW; EEF0754C98740DF6 CRC64;

Query Match 19.9%; Score 179.5; DB 1; Length 168;
Best Local Similarity 30.4%; Pred. No. 2.6e-05;
Matches 52; Conservative 38; Mismatches 72; Indels 9; Gaps 4;

QY 1 MKVKTLSMAILACLIVANSAPFPIGVFNSQSITAMSEAAKAAQKLOSEFG--NEK 57
Db 1 MKFTQFVLTIAINWASAP-AEMKIAVINYQWALLESDAAQYAVDAEKKFGQLNKL 59
```

```
QY 58 TQLEKQAKDLQTKADDDLOAKSAANSQAREDKQREFLELRNFEEKSRDFAIRVEQAENT 117
Db 60 KNLERDAKALQ---DKLVNSGSKMSQGDREKAEIDFKQKARDFOFQSKELNESKAAADR 116

QY 118 LRQYLAEQIYLAETIAKKGLKVLDSAGSGVMYLEKNLDTITKEILEAIN 168
Db 117 MLKKLKPQLDQAVEETIKGGYDVIE--RGAVVDVRPQYDITRQVIERMV 165

RESULT 3
OMPH_YERPS
ID _OMPH_YERPS STANDARD; PRT; 143 AA.
AC P31520;
DT 01-JUL-1993 (Rel. 26, Created)
DT 21-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cationic 19 kDa outer membrane protein precursor (Fragment).
GN OMPH.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=78;
RX MEDLINE=92096452; PubMed=17561172;
RA Vuorio R., Hirvas L., Raybourne R.B., Yu D.T.Y., Vaara M.;
RT "The nucleotide and deduced amino acid sequence of the cationic 19
RT kDa outer membrane protein OmpH of Yersinia pseudotuberculosis."
RL Biochim. Biophys. Acta 1129:124-126(1991).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73247; AAA27657.1; --
DR PIR; S19728; S19728.
DR InterPro; IPR005632; OmpH.
DR Pfam; PF03938; OmpH; 1.
KW Outer membrane; Signal.
FT NON TER 1 1
FT SIGNAL <1 2
FT CHAIN 3 143 CATIONIC 19 kDa OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 143 AA; 16048 MW; E0AC366E62B39A8D CRC64;

Query Match 18.7%; Score 169; DB 1; Length 143;
Best Local Similarity 30.8%; Pred. No. 0.0001;
Matches 44; Conservative 30; Mismatches 63; Indels 6; Gaps 3;

QY 27 IGVFNSQSITAMSEAAKAAQKLOSEFGNEKTQLEKQAKDLQTKADDDLOAKSAANSQAR 86
Db 4 IAIWNVSSIFQQLPAREAVAKQLENEFKGRATELQGMERDLQTKQKLQDQSGTMTKASDR 63

QY 87 EDKQREFLELRNFEEKSRDFA--IRVEQAENTLRQYLAEQIYLAETIAKKGLKVLVD 144
Db 64 TKLENEVMKQRETFSTKAQAFQDNRRQAEE--RNKILSRIDQAVSKVATGGYDVVID 121

QY 145 SAGSGVMYLEKNLDTITKEILEAI 167
Db 122 --ANAVAYADSSKDTADVLKQV 142

RESULT 4
Y157_AQUAE
ID Y157_AQUAE STANDARD; PRT; 162 AA.
AC O66547;
DT 16-OCT-2001 (Rel. 40, Created)
```

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
GN Hypothetical protein AQ_157 precursor.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358 (1998).
CC -!- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M34854; AAA27658.1;
DR InterPro: IPR005632; OmpH.
DR Pfam: PF03938; OmpH; 1.
KW Outer membrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 164
FT SEQUENCE 164 AA; 18225 MW; 70F6C0E5B3C85B7B CRC64;
CC -----
CC Query Match 17.5%; Score 158.5; DB 1; Length 164;
CC Best Local Similarity 26.2%; Pred. No. 0.00057;
CC Matches 45; Conservative 37; Mismatches 61; Indels 16; Gaps 4;
CC -----
QY 15 LVANSFSAFDPPIGVNSQSIAESEAQAQKQLOSEFGNEKTQLEKQAKOLOTQKADDL 74
DB 14 LIAGISFSLDF--ACVDTNKNVRESKFIKAQTELRKE-----LEKYQKLIQEYQKKL 64
QY 75 QA-----KSAAMNQAREDKQRELELRNFEEKSRDFAIRVEQAENTLRQYLAQIYLA 129
DB 65 EALKKSLEKALSEKAKEKEIEQLEDELRLQVEAQSKRKKAELEKMWFDKVIKI 124
QY 130 ATIAKKKGLKLVDSAGSGVMYLEKNLDITKEILEAI 168
DB 125 VESTAKKKIKAVFD--CNSMLYWDKKIDITNEVLKELD 161
CC -----
RESULT 5
OMPH_YEREN
ID OMPH_YEREN STANDARD; PRT; 164 AA.
AC P31519;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cationic 19 kDa outer membrane protein precursor.
GN OMPH.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91123198; PubMed=1991717;
RA Hirvas L., Koski P., Vaara M.;
RT "The ompH gene of Yersinia enterocolitica: cloning, sequencing,
RT expression, and comparison with known enterobacterial ompH
sequences.";
RT J. Bacteriol. 173:1223-1229 (1991).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M34854; AAA27658.1;
DR InterPro: IPR005632; OmpH.
DR Pfam: PF03938; OmpH; 1.
KW Outer membrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 164
FT SEQUENCE 164 AA; 18225 MW; 70F6C0E5B3C85B7B CRC64;
CC -----
CC Query Match 17.5%; Score 158.5; DB 1; Length 164;
CC Best Local Similarity 26.2%; Pred. No. 0.00057;
CC Matches 43; Conservative 36; Mismatches 82; Indels 3; Gaps 2;
CC -----
QY 4 KTLSSMAILACLIVANSFSAFDPPIGVNSQSIAESEAQAQKQLOSEFGNEKTQLEKQ 63
DB 3 KWLCAASLGLALAAASVQA-AKIAIVNVSRIFFQQLPESETVAKOLENEFKGRATELQGM 61
QY 64 AKDLOTKADDLQAKSAAMNQAREDKQRELELRNFEEKSRDFAIRVEQAENTLRQYLA 123
DB 62 ESDLOTKMKQLQDGGSTMKASDRTKLENDVMKQRETFSTYKAQAFQEDNRRRQMEERNKIL 121
QY 124 EQIYLAETIATKKGLKLVDSAGSGVMYLEKNLDITKEILEAI 167
DB 122 SRIQDAKVSVAQSGGYDVVID--ANAVAYADPSKIDTADVLKQV 163
CC -----
RESULT 6
OMPH_VIBCH
ID OMPH_VIBCH STANDARD; PRT; 169 AA.
AC Q9KWL1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein ompH precursor.
GN OMPH OR VC2251.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



RX MEDLINE=88329735; PubMed=2843433;  
 RA Holck A., Kleppe K.;  
 RT "Cloning and sequencing of the gene for the DNA-binding 17K protein  
 of *Escherichia coli*.";  
 RL Gene 67:117-124 (1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12";  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12 / W3110;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,  
 RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;  
 RT "Systematic sequencing of the *Escherichia coli* genome: analysis of the  
 4.0 - 6.0 min (189,987 - 281,416bp) region.";  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
 RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
 RA Lashkari D., Lew H., Lin D., Namath A., Oetner P., Roberts D.,  
 RA Davis R.W.;  
 RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O6:H7 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
 RL Nature 409:529-533 (2001).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayaishi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22 (2001).  
 RN [8]  
 RP SEQUENCE OF 72-161 FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
 RX MEDLINE=91100302; PubMed=1987124;  
 RA Dicker I.B., Seetharam S.R.;  
 RT "Cloning and nucleotide sequence of the *firA* gene and the *firA200* (Ts)  
 allele from *Escherichia coli*.";

J. Bacteriol. 173:334-344 (1991).  
 [9]  
 RN SEQUENCE OF 21-32.  
 RP SPECIES=E.coli; STRAIN=K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.J., Robison K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 in the genome of *Escherichia coli* K-12";  
 RL Electrophoresis 18:1259-1313 (1997).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 through comparison with genomes of *Escherichia coli* K12 and O157";  
 RL Nucleic Acids Res. 30:4432-4441 (2002).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Ferra N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of *Shigella*  
*flexneri* serotype 2a strain 2457T";  
 RL Infect. Immun. 71:2775-2786 (2003).  
 RN [12]  
 RP SIMILARITY TO S-TYPHIMURIUM OMPH.  
 RC SPECIES=E.coli;  
 RX MEDLINE=90201355; PubMed=2318304;  
 RA Hirvas L., Coleman J., Koski P., Vaara M.;  
 RT "Bacterial 'histone-like protein I' (HLP-I) is an outer membrane  
 constituent";  
 RL FEBS Lett. 262:123-126 (1990).  
 CC -1- SUBUNIT: Homotetramer.  
 CC -1- SUBCELLULAR LOCATION: EITHER IN THE NUCLEOID (CHROMATIN) OR IN  
 THE OUTER MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M21118; AAA24630.1; -;  
 DR EMBL; AE000127; AAC73289.1; -;  
 DR EMBL; D83536; BAA77853.1; -;  
 DR EMBL; U70214; AAB08607.1; -;  
 DR EMBL; AE016755; AAN78707.1; -;  
 DR EMBL; AE005194; AAG54480.1; -;  
 DR EMBL; AF002550; BAB33603.1; -;  
 DR EMBL; X54797; CAA38567.1; -;  
 DR EMBL; X75465; CAA53207.1; -;  
 DR EMBL; AE015054; AAN41830.1; -;  
 DR EMBL; AE016978; AAP15711.1; -;  
 DR PIR; D85502; D85502.  
 DR PIR; D90651; D90651.  
 DR PIR; JT0304; DNEC17.  
 DR EcoGene; EGI0455; hlpA.  
 DR InterPro; IPR005632; Omph.  
 DR Pfam; PF03938; Omph; 1.  
 KW DNA-binding; Outer membrane; Signal; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 21 161 HISTONE-LIKE PROTEIN HLP-I.



```

SQ SEQUENCE 161 AA; 17688 MW; 2A966BBD83F3E675 CRC64;
  Query Match 14.9%; Score 134.5; DB 1; Length 161;
  Best Local Similarity 25.8%; Pred. No. 0.02;
  Matches 40; Conservative 36; Mismatches 72; Indels 7; Gaps 4;

QY 14 LLVANSAFSADFFPIGVFNSQSIAESEAAKAAQKQKLOSEFNEKTQLEKQAKDLQTKADD 73
Db 12 LALATSAQAAD-KIAIVNMGSLFQQVAKTGVNTLENEFKGRASELQRMETDLQAKWK 70
QY 74 LQAKSAAMNQAREDKOREFLELRNFEKSRDFAIRVEQAENTLROYLAEOIYLAETI 133
Db 71 LQSMKAGSD---RTKLEKDVMSQRTFAQAQAEKDRARRSRNEERKNLVTRIQTAVK 127
QY 134 AKKGLKLVLDSAGSVVYLEKNL-DITKEILEAI 167
Db 128 ANSQDIDLVD--ANAVYNSDVKDITADVLKQV 160

RESULT 9
OMPH SALTY
ID _OMPH SALTY STANDARD; PRT; 161 AA.
AC P16974;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer membrane protein ompH precursor (Cationic 16 kDa outer membrane
  protein).
DE OMPH OR SPM0225 OR STY0248 OR T0226.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-30.
RC SPECIES=S.typhimurium;
RX MEDLINE=90037020; PubMed=2681205;
RX Koski P., Rhee M., Kantele J., Vaara M.;
RT "Isolation, cloning, and primary structure of a cationic 16-kDa outer
  membrane protein of Salmonella typhimurium."
RL J. Biol. Chem. 264:18973-18981(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=90255961; PubMed=2187745;
RX Koski P., Hirvas L., Vaara M.;
RT "Complete sequence of the ompH gene encoding the 16-kDa cationic
  outer membrane protein of Salmonella typhimurium."
RL Gene 88:117-120(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
  Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
  Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
  Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
  Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
  LT2."
RL Nature 413:852-856(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
  Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
  Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
  Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
  Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jajels K.,
  Krogsh A., Larsen T.S., Leach S., Moule S., O'Gaora P., Parry C.,
  Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
  Whitehead S., Barrell B.G.;
  "Complete genome sequence of a multiple drug resistant Salmonella
  enterica serovar Typhi CT18."
  Nature 413:848-852(2001).
  [5]
  SEQUENCE FROM N.A.
  RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
  RX MEDLINE=22531367; PubMed=12644504;
  RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
  Burland V., Kodymiani V., Schwartz D.C., Blattner F.R.;
  RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
    and CT18."
  J. Bacteriol. 185:2330-2337(2003).
  [6]
  SIMILARITY TO E. COLI HPLA
  RX MEDLINE=90201355; PubMed=2318304;
  RX Hirvas L., Coleman J., Koski P., Vaara M.;
  RT "Bacterial 'histone-like protein I' (HLP-I) is an outer membrane
    constituent?";
  FEBS Lett. 262:123-126(1990).
  CC -1- SUBCELLULAR LOCATION: Outer membrane.
  CC -1- SIMILARITY: BELONGS TO THE OMPH/HLEA FAMILY.
  CC
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@sib-sib.ch).
  CC
  CC EMBL; J05101; AAA27170.1; --
  DR EMBL; AE008705; AAL19189.1; --
  DR EMBL; AL627266; CAD08683.1; --
  DR EMBL; AE016834; AAC67956.1; --
  DR PIR; JQ0528; S09104.
  DR StyGene; SG10265; OmpH.
  DR InterPro; IPR005632; OmpH.
  DR Pfam; PF03938; OmpH; 1.
  KW Outer membrane; Signal; Complete proteome.
  FT SIGNAL 1 20
  FT CHAIN 21 161
  FT SIGNAL 21 161
  SQ SEQUENCE 161 AA; 17905 MW; CF04716C1F7A117D CRC64;
  Query Match 14.5%; Score 131.5; DB 1; Length 161;
  Best Local Similarity 25.2%; Pred. No. 0.031;
  Matches 39; Conservative 35; Mismatches 74; Indels 7; Gaps 4;

QY 14 LLVANSAFSADFFPIGVFNSQSIAESEAAKAAQKQKLOSEFNEKTQLEKQAKDLQTKADD 73
Db 12 LAMVTSQAAD-KIAIVNMGSLFQQVAKTGVNTLENEFKGRASELQRMETDLQAKWK 70
QY 74 LQAKSAAMNQAREDKOREFLELRNFEKSRDFAIRVEQAENTLROYLAEOIYLAETI 133
Db 71 LQSMKAGSD---RTKLEKDVMSQRTFAQAQAEKDRARRSRNEERKNLVTRIQTAVK 127
QY 134 AKKGLKLVLDSAGSVVYLEKNL-DITKEILEAI 167
Db 128 ANDQSIDLVD--ANTVAYNSDVKDITADVLKQV 160

RESULT 10
OMPH CHLPN
ID _OMPH CHLPN STANDARD; PRT; 171 AA.
AC Q9Z8N7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE OMPH-like protein precursor.
GN CPN0301 OR CP0457 OR CPJ0301 OR CPB0310.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=CWL029;
RC  MEDLINE=99206606; PubMed=10192388;
RX  Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA  Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT  "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL  Nat. Genet. 21:385-389(1999).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AR39;
RX  MEDLINE=20150255; PubMed=10684935;
RA  Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA  White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA  Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA  Eisen J., Fraser C.M.;
RT  "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT  pneumoniae AR39.";
RL  Nucleic Acids Res. 28:1397-1406(2000).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=J138;
RX  MEDLINE=20330349; PubMed=10871362;
RA  Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA  Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT  "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT  from Japan and CWL029 from USA.";
RL  Nucleic Acids Res. 28:2311-2314(2000).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=TW-183;
RA  Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
RA  Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT  "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT  other Chlamydia strains based on whole genome sequence analysis.";
RL  Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC  -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC  -!- SIMILARITY: BELONGS TO THE OMPH/HLPA FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AEO01615; AAD18450.1; -
CC  DR  EMBL; AEO02207; AAF38295.1; -
CC  DR  EMBL; AEO02546; BAA98511.1; -
CC  DR  EMBL; AEO17158; AAP98243.1; -
CC  DR  PIR; E72094; E72094.
CC  PIR; E86528; E86528.
CC  DR  PHCI-2DPAGE; Q9Z8N7; -
CC  TIGR; CP0457; -
CC  DR  InterPro; IPR005632; Omph.
CC  DR  Pfam; PF03938; Omph; 1.
CC  KW  Outer membrane; Signal; Complete proteome.
FT  SIGNAL 1 21 POTENTIAL.
FT  CHAIN 22 171 OMPH-LIKE PROTEIN.
SQ  SEQUENCE 171 AA; 19482 MW; E344125CC043CC2F CRC64;

Query Match 14.5%; Score 131; DB 1; Length 171;
Best Local Similarity 22.7%; Pred. No. 0.035;
Matches 37; Conservative 44; Mismatches 76; Indels 6; Gaps 3;

QY 14 LLVANSAGFADFPPIGVNSQSIAMSEAAKQKLOS---EPGNEKTOLEKQAKDLQTK 70
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 LLVLGTSAAHNLGVNLKRCLESDLGKKEELEAKVQQVKNKAEIKELTSTYNNK 69
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 71 ADDLQAKSAMNSQNAEDKQREFLELRNPFEEKSRDFAIRVEQAENTLRQYLAEQTYLAA 130
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 70 LQD-EDYMESLSDSASEELRKKFEDLSGEYNAYQSQYQSQINQSVKRIQKLIQEVKIAA 128
QY 131 ETIAKKGLKLVLDSAGSGVMYLEKNLDITKTEILEINAAWK 173
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 ESVRSKEKLEAILNEE--AVLAIAPTGDKTKTEIATINSEFKK 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
AP04_PIG STANDARD; PRT; 382 AA.
ID AP04_PIG AC O46409;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Apolipoprotein A-IV precursor (Apo-AIV).
GN APOA4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Osada J., Iturralde M., Calleja L., Gonzalez N., Pineiro A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May have a role in chylomicrons and VLDL secretion and
CC catabolism. Required for efficient activation of lipoprotein
CC lipase by Apoc-II; potent activator of LCAT. ApoA-IV is a major
CC component of HDL and chylomicrons.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Secreted in plasma.
CC -!- DOMAIN: Nine of the thirteen 22-amino acid tandem repeats (each
CC 22-mer is actually a tandem array of two A and B, related 11-
CC mers) occurring in this sequence are predicted to be highly alpha-
CC helical, and many of these helices are amphipathic. They may
CC therefore serve as lipid-binding domains with lecithin:cholesterol
CC acyltransferase (LCAT) activating abilities.
CC -!- SIMILARITY: Belongs to the apolipoprotein A1/A4/E family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AJ222966; CAA11020.1; -
CC  DR  HSSP; P32851; 1BR0.
CC  DR  InterPro; IPR000074; Apolipoprotein.
CC  PFam; PF01442; Apolipoprotein; 2.
CC  KW  Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.
FT  SIGNAL 1 20 BY SIMILARITY.
FT  CHAIN 21 382 APOLIPOPROTEIN A-IV.
FT  DOMAIN 33 330 13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT  REPEAT 33 54 1.
FT  REPEAT 60 81 2.
FT  REPEAT 82 103 3.
FT  REPEAT 115 136 4.
FT  REPEAT 137 158 5.
FT  REPEAT 159 180 6.
FT  REPEAT 181 202 7.
FT  REPEAT 203 224 8.
FT  REPEAT 225 246 9.
FT  REPEAT 247 268 10.
FT  REPEAT 269 286 11.
FT  REPEAT 287 308 12.
FT  REPEAT 309 330 13.
FT  DOMAIN 360 377 GLU/GLN-RICH.
SQ  SEQUENCE 382 AA; 43294 MW; 2AC88A02D4379EFC CRC64;

Query Match 13.1%; Score 118.5; DB 1; Length 382;
Best Local Similarity 26.8%; Pred. No. 0.51;
Matches 34; Conservative 29; Mismatches 47; Indels 17; Gaps 4;

```

QY 42 AKAAOKKLOSEFGNEKTOLEKQAKDIQTK-----ADDLQAKSAAMNSQAR-----EDKQ 90  
 Db 227 AQDVQEKLNHQLGEGAFQMKKQAEELKAKISANAEELRQKLVPPVAENVGHGLKNGTEGLQ 286  
 QY 91 REFLELRNFEKSRDFAIRVEQAEINTLRQVLAQIYLAETIAKKKGLKLVLDASGVS 150  
 Db 287 KSULELRSHLDQQVEERFLKVPYGETFNKALVQGV----EDLFQKLG--PLAGDVGHL 340  
 QY 151 MYLEKXN 157  
 Db 341 SFLEKDL 347

## RESULT 12

MFPI\_LYCES  
 ID MFPI\_LYCES STANDARD; PRT; 697 AA.  
 AC P93203;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE MAR binding filament-like protein 1.  
 GN MFPI.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. VENT Cherry; TISSUE=fruit;  
 %RX MEDLINE=97112036; PubMed=8953774;  
 RA Meier I., Phelan T., Gruissem W., Spiker S., Schneider D.;  
 RT "MFPI, a novel plant filament-like protein with affinity for matrix  
 attachment region DNA";  
 RL Plant Cell 8:2105-2115(1996).  
 CC -!- FUNCTION: Binds DNA. Interacts with chromatin via matrix  
 attachment regions (MARs). Likely to participate in nuclear  
 architecture by connecting chromatin with the nuclear matrix and  
 potentially with the nuclear envelope.  
 CC -!- SUBCELLULAR LOCATION: Nuclear matrix.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; Y07861; CAA69181.1; --  
 DR PIR; T07111; T07111.  
 DR Nuclear protein; DNA-binding; Coiled coil.  
 FT DOMAIN 125 681 COILED COIL (POTENTIAL).  
 FT DOMAIN 10 15 POLY-SER.  
 SQ SEQUENCE 697 AA; 79516 MW; 700A56D68D6A7E49 CRC64;

Query Match 12.9%; Score 116.5; DB 1; Length 697;  
 Best Local Similarity 23.8%; Pred. No. 1.2; Indels 41; Gaps 6;  
 Matches 41; Conservative 38; Mismatches 52;  
 QY 33 QSIAMEBAKAAQK-----LOSEFGNEKTOLEKQAKDIQTKA-- 71  
 Db 262 KEVSLRNTSKLAKESEWNSLSWYQSQDQLMNLTSSEIKELKDEIQKRELELKVCS 321  
 QY 72 -DDLQAKSAAMNSQAREK-----QREFLRNFKEKSRDFAIRVEQAEINTLRQVLA 124  
 Db 322 EDNLVQLNSLLERDESKEKELHAIQKEYSFKNSDEKVASDNLGSEQ-EKRLHQ-LEE 379  
 QY 125 QI-----VLAETIAKKKGLKLVLDASGVSVMYLEKXNLDITKEILE 165  
 Db 380 QLGTALEASAKNEVLIALDTREKLRNRWVDLNDVNLKQAEIVTQESLE 431

## RESULT 13

MYH3\_RAT  
 ID MYH3\_RAT STANDARD; PRT; 1940 AA.  
 AC P12847;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, fast skeletal muscle, embryonic.  
 GN MYH3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87060986; PubMed=3783701;  
 RA Strehler E.E., Strehler-Page M.-A., Perriard J.C., Periasamy M.,  
 RA Nadal-Ginard B.;  
 RT "Complete nucleotide and encoded amino acid sequence of a mammalian  
 myosin heavy chain gene. Evidence against intron-dependent evolution  
 of the rod.";  
 RL J. Mol. Biol. 190:291-317(1986).  
 CC -!- FUNCTION: Muscle contraction.  
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 and 2 regulatory light chain subunits (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 characteristic for alpha-helical coiled coils.  
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
 meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be  
 split further into 2 globular subfragments (S1) and 1 rod-shaped  
 subfragment (S2).  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X04267; CAA27817.1; --  
 DR PIR; A24922; A24922.  
 DR HGSP; P13538; 2MYS.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR001609; myosin head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR PRODOM; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS50096; IQ; 1.  
 DR PROSITE; PS50096; IQ; 1.  
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.  
 FT DOMAIN 1 781 MYOSIN HEAD-LIKE.  
 FT DOMAIN 782 811 IQ.  
 FT DOMAIN 840 1933 COILED COIL (POTENTIAL).  
 FT NP\_BIND 179 186 ATP (POTENTIAL).  
 FT DOMAIN 656 678 ACTIN-BINDING.  
 FT DOMAIN 758 772 ACTIN-BINDING.  
 FT MOD\_RES 130 130 METHYLATION (TRI-) (POTENTIAL).  
 SQ SEQUENCE 1940 AA; 223857 MW; B5D546A596ESA696 CRC64;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 09:32:58 ; Search time 39 Seconds  
(without alignments)  
1504.779 Million cell updates/sec

Title: US-10-018-290A-1

Perfect score: 904

Sequence: 1 MKVKTLSMALLLVANSA.....INAAWKGGSKPEMANRXK 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	209	23.1	184	16 Q8XZ12	Q8xz12 ralstonia s
2	187.5	20.7	167	16 Q88MH1	Q88mh1 pseudomonas
3	185.5	20.5	167	16 Q88GN4	Q88gn4 pseudomonas
4	170	18.8	191	2 Q8VL03	Q8vl03 actinobacil
5	166.5	18.4	164	2 Q05360	Q05360 yersinia en
6	163	18.0	165	2 Q9S340	Q9s340 photorhabdu
7	161	17.8	187	16 Q7WJ85	Q7wj85 bordetella
8	161	17.8	187	16 Q7VVC1	Q7vvc1 bordetella
9	161	17.8	203	16 Q7WA51	Q7wa51 bordetella
10	152	16.8	168	16 Q7VRD7	Q7vrd7 candidatus
11	150.5	16.6	187	16 Q82U04	Q82u04 nitrosomona
12	140	15.5	169	16 Q8D2H1	Q8d2h1 wigglewort
13	137	15.2	171	16 Q8A1E2	Q8a1e2 bacteroides
14	136	15.0	195	2 Q9S693	Q9s693 haemophilus
15	135.5	15.0	168	16 Q8EGG6	Q8egg6 shewanella
16	134	14.8	166	16 Q9K1H1	Q9k1h1 reisseria m

17	133.5	14.8	189	2 Q9S6A7	Q9s6a7 haemophilus
18	133.5	14.8	189	2 Q9S6A1	Q9s6a1 haemophilus
19	133.5	14.7	186	16 Q9JX30	Q9jx30 neisseria m
20	132	14.6	186	2 Q9S6A0	Q9s6a0 haemophilus
21	131.5	14.5	182	2 Q9S695	Q9s695 haemophilus
22	131.5	14.5	182	2 Q9S694	Q9s694 haemophilus
23	131.5	14.5	185	2 Q9S698	Q9s698 haemophilus
24	131.5	14.5	186	2 Q9S6A2	Q9s6a2 haemophilus
25	131.5	14.5	189	2 Q9S6A4	Q9s6a4 haemophilus
26	131.5	14.5	193	2 Q9X4F4	Q9x4f4 haemophilus
27	131	14.5	188	2 Q9S6A8	Q9s6a8 haemophilus
28	130.5	14.4	188	2 Q9S6A6	Q9s6a6 haemophilus
29	130.5	14.4	188	2 Q9S6A5	Q9s6a5 haemophilus
30	129.5	14.3	188	2 Q9S6A3	Q9s6a3 haemophilus
31	129	14.3	185	2 Q9S697	Q9s697 haemophilus
32	129	14.3	188	2 Q9S6A9	Q9s6a9 haemophilus
33	128.5	14.2	484	2 Q10372	Q10372 streptococc
34	127.5	14.1	484	16 Q99XV0	Q99xv0 streptococc
35	126	13.9	174	2 Q9S696	Q9s696 haemophilus
36	125.5	13.9	1708	5 Q9U0S6	Q9u0s6 mytilus gal
37	125	13.8	183	16 Q8KFR8	Q8kfr8 chlorobium
38	125	13.8	259	16 Q7VM23	Q7vm23 haemophilus
39	124.5	13.8	227	2 Q10371	Q10371 streptococc
40	124.5	13.8	465	2 Q83XW0	Q83xw0 streptococc
41	123.5	13.7	484	2 Q05464	Q05464 streptococc
42	123	13.6	165	16 Q83DT1	Q83dt1 coxiella bu
43	123	13.6	1941	13 Q8UWA0	Q8uwa0 gallus gall
44	122	13.5	200	2 Q8GKW6	Q8gkw6 streptococc
45	122	13.5	200	2 Q8GKW1	Q8gkw1 streptococc

## ALIGNMENTS

### RESULT 1

Q8XZ12 ID Q8XZ12 PRELIMINARY; PRT; 184 AA.  
AC Q8XZ12;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Probable transmembrane protein.  
GN RSC1413 OR RS05279.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=11681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choise N., Moisan A., Claudel-Renard C., Cunac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Siguier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL; AL646064; CAD15115.1; -.  
DR InterPro; IPR005632; Omph.  
DR Pfam; PF03938; Omph; 1.  
KW Complete proteome.  
SQ SEQUENCE 184 AA; 20765 MW; 6AFA36FEA14FBF48 CRC64;

Query Match 23.1%; Score 209; DB 16; Length 184;  
Best Local Similarity 30.5%; Pred.No.3.1e-07;  
Matches 50; Conservative 37; Mismatches 73; Indels 4; Gaps 1;  
Qy 7 SMATCLLVANSAFSDPFGVNSOSIAMESAAKAAOKKLOSERGNEKTOLEKAKD 66  
Db 21 AFAALAAAFALPATAQEARIAAVNSRILLDSQFAKAAQAKLETFKRDRELQDMAK 80

QY 67 LQTKADLQKSAAMNSQAREDKQRELELRNFEKSRDFAIRVEQAENTLRQYLAEQI 126  
 Db 81 LKMSDKLDKDTAVLADSDTRQRELELRNFEKSRDFAIRVEQAENTLRQYLAEQI 140

QY 127 YLAETIAETIAKKGLKVLDSAGSVMYLEKNDITKEILEAINAA 170  
 Db 141 NRVIRQAEQKYDLVQEA-----VYVNPRIIDITDEVIRKALNAS 180

## RESULT 2

Q88MH1 PRELIMINARY; PRT; 167 AA.  
 AC Q88MH1; 2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Outer membrane protein OmpH.  
 GN OMPH OR PPI1600

OS Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=160488;

QY SEQUENCE FROM N.A.  
 RX MEDLINE=22423060; PubMed=12534463;  
 RA Nelson K.E., Wetzel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
 RA Chris Lee P., Holtzapfe E., Scanlan D., Tran K., Moazzes A.,  
 RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,  
 RA Lauber J., Stjepandic D., Hehseisel J., Straetz M., Heim S.,  
 RA Kewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative analysis of the  
 RT metabolically versatile Pseudomonas putida KT2440.";  
 RL Environ. Microbiol. 4:799-808(2002).  
 DR EMBL; AB016779; AAN67221.1; -.  
 DR TIGR; FPI1600; -.  
 DR InterPro; IPR005632; OmpH.  
 DR Pfam; PF03938; OmpH; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 167 AA; 18809 MW; 02B7DBB4DA3ECF0E CRC64;

Query Match 20.7%; Score 187.5; DB 16; Length 167;  
 Best Local Similarity 31.8%; Pred. No. 8.7e-06;  
 Matches 54; Conservative 35; Mismatches 72; Indels 9; Gaps 4;

QY 4 KTLMAILACLVANSAFSADFFIGVNSQSIAMSEAAKQKQSEFGNEKTQ--L 60  
 Db 3 KLAQVVAALVATPAP-ADMKIYVINYQVALLSDAKKYAVDAEKKFGPQTKLSL 61

QY 61 EKQADLQTKADLQKSAAMNSQAREDKQRELELRNFEKSRDFAIRVEQAENTLRQ 120  
 Db 62 ESSAKGIQ--DRLKGGDKMQQERERLEFQKARDFOQSKELNEAKAVADRMKL 118

QY 121 YLAQYVIAETIAKKGLKVLDSAGSVMYLEKNDITKEILEAINAA 170  
 Db 119 QLKPKLDGAVEVKKGGYDLVLE--RGAVIDVKPQYDITRQVIERMNA 166

## RESULT 3

Q886N4 PRELIMINARY; PRT; 167 AA.  
 AC Q886N4;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Outer membrane protein OmpH, putative.  
 GN PSP01543.

OS Pseudomonas syringae (pv. tomato).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=323;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;

RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
 RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,  
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,  
 RA White O., Fraser C., Collmer A.;  
 RT "Complete sequence of Pseudomonas syringae";  
 RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE016861; AAO55063.1; -.  
 DR TIGR; PSP01543; -.  
 DR InterPro; IPR005632; OmpH.  
 DR Pfam; PF03938; OmpH; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 167 AA; 18899 MW; AE952E29C2763AAA CRC64;

Query Match 20.5%; Score 185.5; DB 16; Length 167;  
 Best Local Similarity 31.0%; Pred. No. 1.2e-05;  
 Matches 52; Conservative 37; Mismatches 70; Indels 9; Gaps 4;

QY 4 KTLMAILACLVANSAFSADFFIGVNSQSIAMSEAAKQKQSEFGNEKTQ--L 60  
 Db 3 KLTQVLVILATVAVSPAP-ADMKIYVINYQVALLSDAKKYAVDAEKKFGPQTKLSL 61

QY 61 EKQADLQTKADLQKSAAMNSQAREDKQRELELRNFEKSRDFAIRVEQAENTLRQ 120  
 Db 62 ESSAKGIQ--DRLVSGGDKMQQERERLEFQKARDFOQSKELNEAKAVADRMKL 118

QY 121 YLAQYVIAETIAKKGLKVLDSAGSVMYLEKNDITKEILEAIN 168  
 Db 119 QLKPKLDGAVEVKKGGFLVFE--RGAVIDVKPQYDITRQVIERMN 164

## RESULT 4

Q8VL03 PRELIMINARY; PRT; 191 AA.  
 AC Q8VL03; 2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Omp1-like protein.  
 GN OMP1.

OS Actinobacillus actinomycetomcomitans (Haemophilus  
 OS actinomycetomcomitans).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Actinobacillus.

QY SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29523;  
 RA Hu W., Teng Y.-T.A.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF321231; AAL55890.1; -.  
 DR InterPro; IPR005632; OmpH.  
 DR Pfam; PF03938; OmpH; 1.  
 SQ SEQUENCE 191 AA; 21491 MW; 2EB12957C1573D02 CRC64;

Query Match 18.8%; Score 170; DB 2; Length 191;  
 Best Local Similarity 28.8%; Pred. No. 0.00017;  
 Matches 55; Conservative 32; Mismatches 80; Indels 24; Gaps 4;

QY 4 KTLMAILACLVANSAFSADFFIGVNSQSIAMSEAAKQKQSEFGNEKTQLEKQ 63  
 Db 3 KIVKLTALSLALAFSSLAMADENIAFYLFQNHDPDKAVAEKLEAEFKPTADKLAEN 62

QY 64 AKDLQTKADLQ--AKSAMSQA-----RE-----DKQRELELRNFE 101  
 Db 63 KKQIDTADIQKKVEAKVAAQKADPKLSADIKKREDEINKYGNQQEINKLIAEHD 122

QY 102 EKSRDFAIRVEQAENTLRQYLAETIAETIAKKGLKVLDSAGSVMYLEKNDITK 161  
 Db 102 EKSRDFAIRVEQAENTLRQYLAETIAETIAKKGLKVLDSAGSVMYLEKNDITK 161

Db 123 OKAKSFOENYAKRENETEKLIVASIQAAATNNVAKQKNYTLVILDDR--SVVYGMGDKNITE 180

QY 162 FILEAINAAWK 172

Db 181 EVLKAIPAQAK 191

RESULT 5

O05360 PRELIMINARY; PRT; 164 AA.

AC O05360;

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Automembrane protein H.

OS Yersinia enterocolitica.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.

OX NCBI\_TaxID=630;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=0:3 Freiburg strain 10543;

RA Lauster R.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: Y12468; CAA73071.1; --

DR InterPro: IPR005632; Omph.

DR Pfam: PF03938; Omph; 1.

SQ SEQUENCE 164 AA; 18305 MW; 31A21F26BDFE317F CRC64;

Query Match 18.4%; Score 166.5; DB 2; Length 164;

Best Local Similarity 27.4%; Pred. No. 0.00025;

Matches 45; Conservative 35; Mismatches 81; Indels 3; Gaps 2;

QY 4 KTLUSMAILACLIVANSAFSDPPIGVFNSSQSIAMSEAAKQKQSEFNGEKTQLEKQ 63

Db 3 KWLCAASLGLAASARVQA-AKTAIVNVSSIFQQLPARETVAKQLENEFKGRATELQGM 61

QY 64 AKDLQTKADDLOAKSAAMNQAREDKQREFLELRNFEESKSRDPAIRVEQAENTLRQYLA 123

Db 62 ERDLQTKQKLDQDGTSMKSSERTNLEKEVMKREFFGKKAQAFQDHRREERNKIL 121

QY 124 EQYLAETIAKKKGLKVLDSAGSVMYLEKNLDTITKEILEAI 167

Db 122 SRIQDAVKSVASKGYDVVID--ANAVAYADPSKDIITADVLKQV 163

RESULT 6

Q9S340

AC Q9S340 PRELIMINARY; PRT; 165 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Periplasmic protein.

GN OMPH.

OS Photorhabdus luminescens (Xenorhabdus luminescens).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Photorhabdus.

OX NCBI\_TaxID=29486;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Hm;

RA Chatonnet-Marton P.I., Givaudan A., Lancois A., Boemare N.E.;

RT "Photorhabdus luminescens genomic region homologous to 4.0 minute

RT Escherichia coli region promotes pleiotropic phenotypes."

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ236920; CAB51930.1; --

DR InterPro: IPR005632; Omph.

DR Pfam: PF03938; Omph; 1.

SQ SEQUENCE 165 AA; 18402 MW; 91BC88EB65FAD767 CRC64;

Query Match

Best Local Similarity 18.0%; Score 163; DB 2; Length 165;

Matches 45; Conservative 35; Mismatches 81; Indels 3; Gaps 2;

Matches 44; Conservative 35; Mismatches 83; Indels 2; Gaps 1;

QY 4 KTLUSMAILACLIVANSAFSDPPIGVFNSSQSIAMSEAAKQKQSEFNGEKTQLEKQ 63

Db 3 KLLCAASFGIALAFSAGAAADKIAVNVYGEIFQQLPAREAVAKQLENEFKGRASELQRM 62

QY 64 AKDLQTKADDLOAKSAAMNQAREDKQREFLELRNFEESKSRDPAIRVEQAENTLRQYLA 123

Db 63 ETDLQSKIQLKLDQDGTSMKSSERTNLEKEVMKREFFGKKAQAFQDHRREERNKIL 122

QY 124 EQYLAETIAKKKGLKVLDSAGSVMYLEKNLDTITKEILEAI 167

Db 123 SRIQDAIKVAGKEGYDIVID--ANAVAYSVSGKNITASVLKQV 164

RESULT 7

Q7WJ85

ID Q7WJ85 PRELIMINARY; PRT; 187 AA.

AC Q7WJ85;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative outer membrane protein.

GN B2614.

OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Bordetella.

OX NCBI\_TaxID=518;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RB50 / ATCC BAA-588;

RA MEDLINE=22827954; PubMed=12910271;

RA Parkhill J., Sebailia M., Preston A., Murphy L.D., Thomson N.,

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,

RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,

RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,

RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagella K.,

RA Leather S., Mouton S., Norbertczak H., O'Neill S., Ormond D., Price C.,

RA Rabbittowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,

RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;

RT "Comparative analysis of the genome sequences of Bordetella pertussis,

RT Bordetella parapertussis and Bordetella bronchiseptica.";

RL Nat. Genet. 35:32-40(2003).

DR EMBL: BX640444; CAB33107.1; --

KW Complete proteome.

SQ SEQUENCE 187 AA; 20933 MW; A6C7F0C70785D0A7 CRC64;

Query Match

Best Local Similarity 17.8%; Score 161; DB 16; Length 187;

Matches 42; Conservative 41; Mismatches 75; Indels 8; Gaps 2;

QY 6 LSMAILACLIVANSAF---SADPPIGVFNSSQSIAMSEAAKQKQSEFNGEKTQLE 61

Db 24 VSLALAGALLFGSSAAVTAQAQCTKIGFVNTIRLRESGPAKAAQKSEFNRDDELQ 83

QY 62 KQAKDLQTKADDLOAKSAAMNQAREDKQREFLELRNFEESKSRDPAIRVEQAENTLRQY 121

Db 84 RLSSLSRSQAQKFDKAPVLSESDRVKQRELNLDMDLQKRRFQEDFNRRNRBFSS 143

QY 122 LAEQIVLAETIAKKKGLKVLDSAGSVMYLEKNLDTITKEILEAI 167

Db 144 IVTKANDAIXIAEQENYDLIIQDA---VTVPRIIDITDKVIQSL 185

RESULT 8

Q7VYCI

ID Q7VYCI PRELIMINARY; PRT; 187 AA.

AC Q7VYCI;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)



```

DE Putative outer membrane protein.
GN B2L428.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP
RC SEQUENCE FROM N.A.
RX STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Rabinowitsch E., Rutter S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Saunders S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640415; CAB41718.1; -.
KW Complete proteome.
SQ SEQUENCE 187 AA; 20933 MW; A6C7F0C70785D0A7 CRC64;

Query Match 17.8%; Score 161; DB 16; Length 187;
Best Local Similarity 25.3%; Pred. No. 0.00071;
Matches 42; Conservative 41; Mismatches 75; Indels 8; Gaps 2;

Qy 6 LSMALACLLVANSF----SADPPIGVNSOSIAMSEAAKAAKKLQSEFGNEKTQLE 61
Db 24 VSLALAGALLFGSSAAVTAQAQGTGKIGFVNTERILRSGPAKAAQSKIESFEKRRDDELQ 83
Qy 62 KQAKDLQTKADDLQAKSAAMNQAREDKQREFFLELRNFEEKSRDFAIRVEQAENTLRQ 121
Db 84 RLSSLSRAQKFKDKAPVLSESDRVKQRELSNLDLQKRRFEQDFNRRNEEFSS 143
Qy 122 LAEQIYLAETIAKKGLKVLDSASGVMYLEKNLDTITKEILEAI 167
Db 144 IVTKANDAIAKRAEQENYDLIIQDA---VTNPRIDITDKVIQSL 185

RESULT 9
Q7WA51
ID Q7WA51 PRELIMINARY; PRT; 203 AA.
AC Q7WA51;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Putative outer membrane protein.
GN BPI1536.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP
RC SEQUENCE FROM N.A.
RX STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Rabinowitsch E., Rutter S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Saunders S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";

```

```

RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640427; CAB36838.1; -.
KW Complete proteome.
SQ SEQUENCE 203 AA; 22824 MW; 596B9D44EF081D8 CRC64;

Query Match 17.8%; Score 161; DB 16; Length 203;
Best Local Similarity 25.3%; Pred. No. 0.00078;
Matches 42; Conservative 41; Mismatches 75; Indels 8; Gaps 2;

Qy 6 LSMALACLLVANSF----SADPPIGVNSOSIAMSEAAKAAKKLQSEFGNEKTQLE 61
Db 40 VSLALAGALLFGSSAAVTAQAQGTGKIGFVNTERILRSGPAKAAQSKIESFEKRRDDELQ 99
Qy 62 KQAKDLQTKADDLQAKSAAMNQAREDKQREFFLELRNFEEKSRDFAIRVEQAENTLRQ 121
Db 100 RLSSLSRAQKFKDKAPVLSESDRVKQRELSNLDLQKRRFEQDFNRRNEEFSS 159
Qy 122 LAEQIYLAETIAKKGLKVLDSASGVMYLEKNLDTITKEILEAI 167
Db 160 IVTKANDAIAKRAEQENYDLIIQDA---VTNPRIDITDKVIQSL 201

RESULT 10
Q7VRD7
ID Q7VRD7 PRELIMINARY; PRT; 168 AA.
AC Q7VRD7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Histone-like protein, located in outer membrane.
GN HLP4 OR BFL280.
OS Candidatus Blochmannia floridanus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX NCBI_TaxID=203907;
RN [1]
RP
RC SEQUENCE FROM N.A.
RX MEDLINE=22784745; PubMed=12886019;
RA Gil R., Silva F.J., Zientz E., Belmotte F., Gonzalez-Candelas F.,
RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
RA van Ham R.C.H.J., Gross R., Moya A.;
RT "The genome sequence of Blochmannia floridanus: comparative analysis
RT of reduced genomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
DR EMBL; BX248585; CAB83351.1; -.
KW Complete proteome.
SQ SEQUENCE 168 AA; 19664 MW; 37BC6B37A5AAB43 CRC64;

Query Match 16.8%; Score 152; DB 16; Length 168;
Best Local Similarity 29.0%; Pred. No. 0.0027;
Matches 49; Conservative 29; Mismatches 77; Indels 14; Gaps 5;

Qy 1 MKVKTLSMAILACLLVANSFADPPIGVNSOSIAMSEAAKAAKKLQSEFGNEKTOL 60
Db 1 MKNRVYMLGIIWLAQINYYNAAN-KIVVMVNAVIFQSSAQRTBIKQLEYEKDRAEL 59
Qy 61 EKQAKDLQTKADDLQAKSAAMNQAREDKQREFFLELRNFEEKSRDFAIRVEQAENTLRQ 120
Db 60 EMMEHDLQTKMQLQRDGATMTASDRNTLEKSLIAQELFSNKAFL-----QOENHARQ 114

RESULT 11
Q82U04
ID Q82U04 PRELIMINARY; PRT; 187 AA.
AC Q82U04;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative transmembrane protein.

```

Query Match	15.2%	Score 137;	DB 16;	Length 171;
Best local Similarity	25.1%	Pred. No. 0.03;		
Matches	44;	Conservative 44;	Mismatches 71;	Indels 16; Gaps 5
QY	1	MKVKTLSMAILACLLVANSAFSDPFGVFNFSQSIAMSESAAKAAKKLQSEFGNKTQL	60	
Db	1	MRKSVLSIMLLFAISMAASQK-----FALIDTEVILKNIPIYQSANEOQLQATKKYQSEV	56	
QY	61	EKAQDLQTKRADDLQAKSAAMS-----NOARED-----KOREFLELRNFPEKSRDRFAIRVEQ	113	
Db	57	EVIAKEAQKMFQDYQAOSSTLSAAQTKKEDEIVAKESAAELKKYKPGSELAKMQEK	116	
QY	114	AENTLROYLAEQIYLAETIAKKGLKLVLDSAGS--VMYLEKNLDTIKTILEAI	167	
Db	117	LINPIQ---DEIYGAVKELSQHLGYDLVLDRASAAGIIFANPRIDISDEVLRKL	167	
RESULT 14				
Q9S693		PRELIMINARY;	PRT;	195 AA.
ID	Q9S693			
AC	Q9S693;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Outer membrane protein 26 (Fragment).			
GN	SKP.			
OS	Haemophilus influenzae;			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
CC	NCBI TaxID=727;			



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 09:31:28 ; Search time 59 Seconds  
(without alignments)  
890.743 Million cell updates/sec

Title: US-10-018-290A-1

Perfect score: 904

Sequence: 1 MKVKTLMAILACLLVANS.....INAAKKGGSKLPEMANRKK 186

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	893.5	98.8	187	4	AAB37887 Lawsonia
2	159	17.6	166	6	ABM67564 Photorhab
3	134.5	14.9	197	2	AAW10221 Haemophil
4	134	14.8	166	3	AAW75115 Neisseria
5	133	14.7	166	3	AAW75116 Neisseria
6	131	14.5	166	3	AAW75114 Neisseria
7	131	14.5	171	4	AAE04724 Chlamydia
8	131	14.5	171	5	ABB94317 Chlamydia
9	131	14.5	171	5	ABB90564 Chlamydia
10	131	14.5	185	6	ABP80503 N. gonorr
11	131	14.5	185	6	ABP79763 N. gonorr
12	127.5	14.1	484	5	ABP30015 Streptoco
13	118.5	13.1	382	3	ABP90663 Human sec
14	118	13.1	158	2	AAW34897 Chlamydia
15	118	13.1	717	3	AAW21231 Tonato Le
16	116	12.8	177	2	AAW32170 Chlamydia
17	115.5	12.8	170	2	AAW34497 Porphyrom
18	115.5	12.8	174	2	AAW34496 Porphyrom
19	115.5	12.8	177	2	AAW34370 Porphyrom
20	113	12.5	195	3	AAW21229 Tonato E-
21	113	12.5	1957	6	ABM68683 Photorhab
22	112.5	12.4	876	3	AAW77573 Human cyt
23	112.5	12.4	876	4	AAU28104 Novel hum
24	112	12.4	173	5	ABB94352 Chlamydia
25	112	12.4	173	7	ADD43762 Chlamydia

## ALIGNMENTS

## RESULT 1

AAB37887  
 ID AAB37887 standard; protein; 187 AA.  
 XX AC  
 XX AAB37887;  
 XX AC  
 DT 06-MAR-2001 (first entry)  
 XX  
 DE Lawsonia intracellularis OmpH.  
 XX  
 KW Lawsonia intracellularis; OmpH; immunogen; antigen; vaccine;  
 KW avian infection; porcine infection; porcine proliferative enteropathy;  
 KW PPE.  
 XX  
 OS Lawsonia intracellularis.  
 XX  
 PN WO200069905-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 11-MAY-2000; 2000WO-AU000438.  
 XX  
 PR 13-MAY-1999; 99US-0133986P.  
 XX  
 PA (PIGR-) PIG RES & DEV CORP.  
 XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 PI  
 PI Hasse D, Panaccio M, Sinistaj M;  
 XX WPI; 2001-016211/02.  
 DR N-PSDB; AAC88904.  
 XX  
 PT New isolated Lawsonia spp. OmpH polypeptides and nucleic acids, useful  
 for the prophylaxis, treatment and detection of Lawsonia infections.  
 PS  
 PS Claim 6; Page 81-82; 85pp; English.  
 CC  
 CC The present sequence is Lawsonia intracellularis OmpH. Immunogenic OmpH  
 peptides and polypeptides are useful as antigens in a vaccine preparation  
 for conferring humoral immunity against Lawsonia intracellularis and  
 related pathogens in animal hosts. The polypeptides are capable of  
 eliciting the production of antibodies against Lawsonia sp. when  
 administered to an avian or porcine animal. The OmpH polypeptides and  
 polynucleotides can be used for the detection, prophylaxis and treatment  
 of an infection of an animal by Lawsonia sp. They are used particularly  
 for porcine proliferative enteropathy (PPE) infections  
 XX  
 SQ Sequence 187 AA;

AAY36954 Protein w  
 AAY77955 A. thalia  
 AAM40027 Human pol  
 AAB92515 Human pro  
 AAB93006 Human pro  
 ABU08487 S. pneumo  
 AAB47278 PN7771. 8  
 AAO22445 PN7771 pr  
 AAR57365 K39 polyp  
 AAW03691 Leishmani  
 ADE63514 Rat Prote  
 ADE63518 Rat Prote  
 ABM69174 Photorhab  
 ABU03037 Moss lipi  
 ABB71125 Drosophil  
 AAB21233 Corn MFP1  
 ABR53116 Protein s  
 ABP73482 Candida a  
 AAB41592 Human ORF  
 AAG31272 Arabidops

Query Match 98.8%; Score 893.5; DB 4; Length 187;  
 Best Local Similarity 99.5%; Pred. NO. 6.5e-72;  
 Matches 186; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MKVKTLMAILACLLVANSFAFSDFFIGVNSQSIAESEAQAQKQLQSEFGNEKTOL 60  
 Db 1 MKVKTLMAILACLLVANSFAFSDFFIGVNSQSIAESEAQAQKQLQSEFGNEKTOL 60

QY 61 EKOAKDLQKADLQAKSAAMNQAREDKOREFLELRNFEKSRDPFAIRVQAEINTLRQ 120  
 Db 61 EKOAKDLQKADLQAKSAAMNQAREDKOREFLELRNFEKSRDPFAIRVQAEINTLRQ 120

QY 121 YLAEOIYLAETIAKKGKGLVLDASGVSVMYLEKNLDTITKEILEIN-AAWKKGGSKLP 179  
 Db 121 YLAEOIYLAETIAKKGKGLVLDASGVSVMYLEKNLDTITKEILEIN-AAWKKGGSKLP 180

QY 180 EMANRKK 186  
 Db 181 EMANRKK 187

RESULT 2  
 ABM67564  
 ID ABM67564 standard; protein; 166 AA.  
 XX AC ABM67564;  
 XX XX  
 XX 20-NOV-2003 (first entry)  
 XX XX  
 XX Photorhabdus luminescens protein sequence #661.  
 XX XX  
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX XX  
 XX Photorhabdus luminescens.  
 XX XX  
 XX WO200294867-A2.  
 XX XX  
 XX 28-NOV-2002.  
 XX XX  
 XX 07-FEB-2002; 2002WO-IB003040.  
 XX XX  
 XX 07-FEB-2001; 2001FR-00001659.  
 XX XX  
 XX (INSP ) INST PASTEUR.  
 XX XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX XX  
 XX Duchaud E, Taourit S, Glaser P, Prangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 PI XX  
 XX WPI; 2003-148459/14.  
 XX XX  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX XX  
 XX Claim 2; SEQ ID NO 661; 1205pp; French.

The invention relates to the isolation of genes and their encoded proteins from Photorhabdus luminescens. The isolated sequences are sources of probes and primers for detecting the genome of P. luminescens and related species; to study polymorphisms; for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than P. luminescens and are able to alter response or sensitivity to toxins and antibiotics produced by P. luminescens. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically, (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX SQ Sequence 166 AA;

Query Match 17.6%; Score 159; DB 6; Length 166;  
 Best Local Similarity 26.8%; Pred. NO. 2.5e-06;  
 Matches 44; Conservative 34; Mismatches 84; Indels 2; Gaps 1;

QY 4 KTLMAILACLLVANSFAFSDFFIGVNSQSIAESEAQAQKQLQSEFGNEKTOL 63  
 Db 3 KULCAASFGIALAFSVGAQAADKIAVNVGEIFQQLPAREAVVKQLENEFKRASELQRM 62

QY 64 AKDLQTKADDLQAKSAAMNQAREDKOREFLELRNFEKSRDPFAIRVQAEINTLRQ 123  
 Db 63 ETDLQSKIQLQKLDGSGTMSSERTNLEKVMKREFAKKAQAFEQDHRREMEERNKIL 122

QY 124 EQIYLAETIAKKGKGLVLDASGVSVMYLEKNLDTITKEILEAI 167  
 Db 123 SRIQDAIKVAVGEGYDVVID--ANAVAYSVSGKNITTSVLKQV 164

RESULT 3  
 AAW10221  
 ID AAW10221 standard; protein; 197 AA.  
 XX AC AAW10221;  
 XX XX  
 XX 02-JUL-1997 (first entry)  
 XX XX  
 XX Haemophilus influenzae outer membrane protein OMP26.  
 XX XX  
 XX Vaccine; outer membrane protein; OMP26; respiratory tract; otitis media.  
 XX XX  
 XX Haemophilus influenzae.  
 XX XX  
 XX Key Location/Qualifiers  
 XX Peptide 1..23  
 XX /label= signal  
 XX Protein 24..197  
 XX /label= OMP26  
 XX XX  
 XX WO9701638-A1.  
 XX XX  
 XX 16-JAN-1997.  
 XX XX  
 XX 27-JUN-1996; 96WO-GB001549.  
 XX XX  
 XX 27-JUN-1995; 95GB-00013074.  
 XX XX  
 XX (CORT-) CORTECS INT LTD.  
 XX XX  
 XX Kyd J, Cripps A, Smith CJ;  
 XX XX  
 XX WPI; 1997-100209/09.  
 XX XX  
 XX N-ESDB; AAT50959.  
 XX XX  
 XX New antigenic outer membrane protein from Haemophilus influenzae - useful  
 XX in vaccines and as diagnostic reagent.  
 XX Claim 2; Fig 1; 34pp; English.

The present sequence represents the novel protein, outer membrane protein OMP26 from Haemophilus influenzae. This new antigenic protein can be used as an immunogen in vaccines for the treatment or prevention of H. influenzae infections (of the respiratory tract or otitis media), and as a diagnostic reagent for diagnosing such infections. OMP26 if isolated



PT Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics.  
XX  
PS Claim 2; Page 877; 1453pp; English.  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941  
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of the  
CC invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the manufacture  
CC of medicaments for treating or preventing infection due to *Neisseria*  
CC bacteria (e.g. meningitis and septicaemia), to detect the presence of  
CC *Neisseria* bacteria, or to raise antibodies. They may also be used to  
CC screen for agonists or antagonists, which may themselves have use as  
CC antibacterial agents. The polynucleotides of the invention may also be  
CC used in gene therapy protocols  
XX  
SQ Sequence 166 AA;  
  
Query Match 14.7%; Score 133; DB 3; Length 166;  
Best Local Similarity 25.9%; Pred. No. 0.00052;  
Matches 48; Conservative 32; Mismatches 61; Indels 44; Gaps 7;  
  
QY 4 KTLMAILACLLVANSAFSADFPPIGVENSQSIAMSEAAKAAQKLOSEFGNEKTOLEKQ 63  
DB 6 RAFAAALIGLCCTAGAHADTFQKIGFINTERIYLESQAKRIQTLDSFASARQDELQK- 64  
QY 64 AKLQTKADLDLOAKSA-----AMSNQARE-----DKREFLE---LRNFEEKS 104  
DB 65 ---LQREGLDLERQLAEGKLDKAKQAQAEKWCGLVAAPRKKQAQFEEDYNLRN----- 116  
QY 105 RDEAIRVEQAENTLROYLAEGIYLAETIAKKKGLKVLDSASGVMYLEKNLDTITKEIL 164  
DB 117 EEFASLQONANRIVK-----IAKQEGYDVILQ----DVIYVNTQYDVTDSVI 160  
QY 165 EAINA 169  
DB 161 KENNA 165  
  
RESULT 6  
AAZ75114  
ID AAZ75114 standard; protein; 166 AA.  
XX  
AC AAZ75114;  
XX  
DT 12-SEP-2003 (revised)  
DT 21-MAR-2000 (first entry)  
XX  
DE *Neisseria gonorrhoeae* ORF 570 protein sequence SEQ ID NO:1702.  
XX  
KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
KW antibacterial; gene therapy.  
XX  
OS *Neisseria gonorrhoeae*.  
XX  
FN WO9957280-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 30-APR-1999; 99WO-US009346.  
XX  
PR 01-MAY-1998; 98US-0083758P.  
PR 31-JUL-1998; 98US-0094869P.  
PR 02-SEP-1998; 98US-0098994P.  
PR 02-SEP-1998; 98US-0098994P.  
PR 09-OCT-1998; 98US-0103749P.  
PR 09-OCT-1998; 98US-0103749P.  
PR 09-OCT-1998; 98US-0103749P.  
PR 09-OCT-1998; 98US-0103749P.  
PR 25-FEB-1999; 99US-0121528P.

XX  
PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;  
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI; 2000-062150/05.  
DR N-PSDB; AAZ53876.  
XX  
PT Novel *Neisseria* polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics.  
XX  
PS Claim 2; Page 876; 1453pp; English.  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941  
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of the  
CC invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the manufacture  
CC of medicaments for treating or preventing infection due to *Neisseria*  
CC bacteria (e.g. meningitis and septicaemia), to detect the presence of  
CC *Neisseria* bacteria, or to raise antibodies. They may also be used to  
CC screen for agonists or antagonists, which may themselves have use as  
CC antibacterial agents. The polynucleotides of the invention may also be  
CC used in gene therapy protocols. (Updated on 12-SEP-2003 to standardise OS  
CC field)  
XX  
SQ Sequence 166 AA;  
  
Query Match 14.5%; Score 131; DB 3; Length 166;  
Best Local Similarity 29.8%; Pred. No. 0.00079;  
Matches 53; Conservative 34; Mismatches 61; Indels 30; Gaps 10;  
  
QY 6 LSMAILACL--LVANSAFSADF--PIGVENSQSIAMSEAAKAAQKLOSEFGNEKTOLE 61  
DB 4 LTRAFPAALIGLCCTTGAHADTFQKIGFINTERIYLESQAKRIQTLDSFASARQDELQ 63  
QY 62 KQAKDLQTKADLDLOAKSA-----AMSNQAREDKREFLELR---NFEEKSRDPAIRV 111  
DB 64 K---LQREGLDLERQLAEGKLDKAKQAQAEKWCGLVEAFRKKQAQFE---DYNLRR 115  
QY 112 EQAENTLROYLAEGIYLAETIAKKKGLKVLDSASGVMYLEKNLDTITKEIL EAINA 169  
DB 116 NEEFASLQONANRIVV---KIAKQEGYDVILQ----DVIYVNTQYDVTDSVIKENNA 165  
  
RESULT 7  
AAE04724  
ID AAE04724 standard; protein; 171 AA.  
XX  
AC AAE04724;  
XX  
DT 11-SEP-2003 (revised)  
DT 10-SEP-2001 (first entry)  
XX  
DE *Chlamydia pneumoniae* outer membrane protein.  
XX  
KW Outer membrane protein; therapy; *Chlamydia* infection; antibiotic;  
KW vaccine.  
XX  
OS *Chlamydia pneumoniae*.  
XX  
FN WO200146225-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 20-DEC-2000; 2000WO-CA001535.  
XX  
PR 22-DEC-1999; 99US-0171539P.





PR 03-JUL-2000; 2000GB-00016363.  
 PR 11-JUL-2000; 2000GB-00017047.  
 PR 21-JUL-2000; 2000GB-00017983.  
 PR 07-AUG-2000; 2000GB-00019368.  
 PR 19-AUG-2000; 2000GB-00020440.  
 PR 14-SEP-2000; 2000GB-00022583.  
 PR 10-NOV-2000; 2000GB-00027549.  
 PR 22-DEC-2000; 2000GB-00031706.  
 XX (CHIR-) CHIRON SPA.  
 XX  
 XX Ratti G, Grandi G;  
 PI  
 XX  
 DR WPI; 2002-154726/20.  
 DR N-PSDB; ABL91222.  
 XX  
 XX Novel Chlamydia pneumoniae protein useful in the manufacture of a  
 PT medicament for treatment or prevention of infection due to Chlamydia,  
 PT preferably Chlamydia pneumoniae, and for diagnostic purposes.  
 XX  
 PS Claim 1; Page 78; 364pp; English.  
 XX  
 CC Sequences AB90526-AB90715 represent novel proteins from Chlamydia  
 CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding  
 CC them. The proteins are predicted to be immunogenic and may therefore be  
 CC useful in vaccine production and for diagnostic purposes. Chlamydia  
 CC pneumoniae is a common cause of respiratory disease in humans, and is  
 CC also involved in the development of cardiovascular diseases such as  
 CC atherosclerosis, coronary artery disease, carotid artery stenosis,  
 CC myocardial infarction, cerebrovascular disease, aortic aneurysm,  
 CC claudication and stroke. The proteins and nucleic acids of the invention  
 CC may be used in vaccines and pharmaceutical compositions for the  
 CC prevention or treatment of chlamydial infections, particularly Chlamydia  
 CC pneumoniae infections. The proteins may also be used in the detection of  
 CC Chlamydia pneumoniae, and the nucleic acids may be used in the detection of  
 CC Chlamydia pneumoniae gene expression. The present sequence represents a  
 CC specifically claimed Chlamydia pneumoniae protein of the invention.  
 CC (Updated on 29-AUG-2003 to standardise OS field)  
 XX  
 XX Sequence 171 AA;  
 SQ  
 Query Match 14.5%; Score 131; DB 5; Length 171;  
 Best Local Similarity 22.7%; Pred. No. 0.00081;  
 Matches 37; Conservative 44; Mismatches 76; Indels 6; Gaps 3;  
 QY 14 LLVANSAFSADFFPIGVNSQSIAESEAQAQKQLQS---BFGNEKTQLEKQAKDLQTK 70  
 Db 10 LLVLGSTSAAHNIGYVNLKRCLEESDLGKETELEAMKQOFVYKNAEKIEELTSTYNK 69  
 QY 71 ADLQAKSAAMNQAREDKQREFLELRNPEKSRDFAIRVEQAENTLROVLAQIYLAA 130  
 Db 70 LQD-EDYWSLSDSASBELRKKFEDLSGEYNAYOSYQYSINQSNVRIQKLIQEVKIAA 128  
 QY 131 ETIAKKKGLKIVLDSAGSVNMLEKKNLDITKEILEAINAAWK 173  
 Db 129 ESVRSKEKLEAILNEE--AVLAIAPTGTDKTTETIAILNESFKK 169  
 RESULT 10  
 ID ABP80503  
 XX  
 AC ABP80503;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae amino acid sequence SEQ ID 7536.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX

PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 XX 12-FEB-2002; 2002WO-IB002069.  
 PF  
 XX 12-FEB-2001; 2001GB-00003424.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA  
 XX Fontana MR, Pizza M, Masignani V, Monaci E;  
 PI  
 XX WPI; 2003-058415/05.  
 DR N-PSDB; ABZ41473.  
 XX  
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection.  
 PT  
 XX Disclosure; Page 737; 815pp; English.  
 PS  
 XX The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention  
 XX  
 XX Sequence 185 AA;  
 SQ  
 Query Match 14.5%; Score 131; DB 6; Length 185;  
 Best Local Similarity 29.8%; Pred. No. 0.0009;  
 Matches 53; Conservative 34; Mismatches 61; Indels 30; Gaps 10;  
 QY 6 LSMAILACL--LVANSAFSADF--PIGVNSQSIAESEAQAQKQLQSEFGNEKTQLE 61  
 Db 23 LTRAFAAALIGLCCTTGAAHDTFQKIGFINTERIYLESKQARNIQKTLDEGFSARQDELQ 82  
 QY 62 QQAKDLQTKADDLQAKSA-----AMSNQAREDKQREFLELR-----NFEKSRDFAIRV 111  
 Db 83 K----LQREGLDLERQLAGGLKDKAKQAQ--EKKWRGLVEAFKKQKQFEE---DYNLRR 134  
 QY 112 EQAENTLROVLAQIYLAAETIAKKKGLKIVLDSAGSVNMLEKKNLDITKEILEAINA 169  
 Db 135 NEEFASLQQN-ANRVIV---KIAQEGYDVILQ----DVIYVNTQYDVTDSVIKEMNA 184  
 RESULT 11  
 ID ABP79763  
 XX  
 AC ABP79763;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae amino acid sequence SEQ ID 6056.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 XX 12-FEB-2002; 2002WO-IB002069.  
 PF  
 XX 12-FEB-2001; 2001GB-00003424.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA  
 XX Fontana MR, Pizza M, Masignani V, Monaci E;  
 PI

XX WPI: 2003-058415/05.  
 DR N-PSDB; ABZ40733.  
 XX  
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection.  
 XX  
 PS Disclosure; Page 629; 815pp; English.  
 XX  
 CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention  
 XX  
 SQ Sequence 185 AA;  
 Query Match 14.5%; Score 131; DB 6; Length 185;  
 Best Local Similarity 29.8%; Pred. No. 0.0009;  
 Matches 53; Conservative 3; Mismatches 61; Indels 30; Gaps 10;  
 QY 6 LSNAILACL--LVNAGAFSADF--PIGVNSQSIAMESEAAKQKLOQSEFGNEKTQLE 61  
 Db 23 LTRAFAAALIGLCCTTGAAHDTFKIGFINTERIYLESKQARNIQKTLGDFSPARQDELQ 82  
 QY 62 KQAKDLQTKADDLOAKSA-----AMSNQAREDKQREFLELR-----NFEKSRDPAIRV 111  
 Db 83 K-----LQREGLDLERQLAGGKLDAKKAQA--BEKWRGLVEAFRKKQAQFE---DYNLRR 134  
 QY 112 EQAENTLRQVLAEQIYLAETIAKKKGLKLVLDSSAGSVMYLEKNLDITKEILEATNA 169  
 Db 135 NEEFASLQQN-ANRVIV---KIAKQEGYDVLQ-----DVIYVNTQYDVTSDVIKEMNA 184  
 -RESULT 12  
 ABP30015  
 ID ABP30015 standard; protein; 484 AA.  
 XX  
 AC ABP30015;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 9206.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB004789.  
 XX  
 PR 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Maignani V, Margarit Y RosI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 XX WPI: 2002-352536/38.  
 DR N-PSDB; ABN70646.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection or

PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX  
 PS Claim 1; Page 4044; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B.  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX  
 SQ Sequence 484 AA;  
 Query Match 14.1%; Score 127.5; DB 5; Length 484;  
 Best Local Similarity 25.3%; Pred. No. 0.006;  
 Matches 49; Conservative 27; Mismatches 77; Indels 41; Gaps 6;  
 QY 2 KVTLSMAILACILLVANSAFSADFP1---GVFNSQSIAMESEAAKQKLOQSEFGNE-- 56  
 Db 14 KLKTGTASVAVALTVLGAGFANQTEVKANGDGNPREVIEDLAANNPAIQVIRLYENKDL 73  
 QY 57 -----KTQLEKQAKDLQTKADDLOAKSAAMSNQAREDKQRE 92  
 Db 74 KARLENAMVAGRDFKRAEELEKAKQALEQDKLQLEKQAKQLEKQAKQLEKQAKQLEKQAKQ 132  
 QY 93 FLELRNFEEKSRDFAIRVEQAEQTLRQVLAEQIYLAETIAKKKGLKLVLDSSA---GS 149  
 Db 133 ---LEKELEKKEALELAIDQASRDYHRATA---LEKELEKKEKALELAIDQASQDYNR 185  
 QY 150 VMYLEKNLD-ITKE 162  
 Db 186 ANVLEKELETTIRE 199  
 RESULT 13  
 AAB90663  
 ID AAB90663 standard; protein; 382 AA.  
 XX  
 AC AAB90663;  
 XX  
 DT 01-JUN-2001 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 206.  
 XX  
 KW Human; secreted protein; immunomodulatory; antisclerotic; dermatological;  
 KW antiinflammatory; anti-HIV; cytostatic; cardiant; vascular;  
 KW anti-angiogenic; ophthalmological; neuroprotectant; nootropic;  
 KW anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial;  
 KW vulnery; vaccine; gene therapy; cancer; protein coordinate data;  
 XX infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200121658-A1.  
 XX  
 PD 29-MAR-2001.  
 XX  
 XX 22-SEP-2000; 2000WO-US026013.  
 XX  
 XX 24-SEP-1999; 99US-0155709P.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;  
 XX WPI; 2001-235311/24.  
 DR  
 XX Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease  
 PT and diabetic retinopathy.  
 XX  
 XX Disclosure; Page 869-870; 890pp; English.  
 PS  
 XX The present sequence is provided in a specification relating to nucleic  
 CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic  
 CC acid molecules and polypeptides may be used in the prevention, diagnosis  
 CC and treatment of diseases such as immune disorders (e.g. multiple  
 CC sclerosis, systemic lupus erythematosus and human immuno-deficiency virus  
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and  
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,  
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic  
 CC disorders (e.g. corneal graft neovascularisation and diabetic  
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,  
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or  
 CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic  
 CC acid molecules may be used to produce the secreted polypeptides. They may  
 CC also be used as DNA probes in diagnostic assays to detect and quantitate  
 CC the presence of similar nucleic acid sequences in samples. The  
 CC polypeptides may be used as antigens in the production of antibodies and  
 CC in assays to identify modulators of their expression and activity  
 XX  
 SQ Sequence 382 AA;  
 Query Match 13.1%; Score 118.5; DB 4; Length 382;  
 Best Local Similarity 26.8%; Pred. No. 0.029;  
 Matches 34; Conservative 29; Mismatches 47; Indels 17; Gaps 4;  
 QY 42 AKAAQKLOSEFGNEKTKLEKQAKDLQTK----ADDLQAKSAAMSNOAR-----EDKQ 90  
 DB 227 AQDVQEKLNHQLGLEAFQMKKQABELKAKISANADELQKLVPAENVNHGHLKGNTEGLQ 286  
 QY 91 REFLELRNFEKSRDFAIRVEQAENTLRQVLAETIAAKKGLKLVLDSSAGSV 150  
 DB 287 KSJLELRSHLDQQVEERFLKVPYGETFNKALVQGV---EDLRQKLG--PLAGDVEGHL 340  
 QY 151 MYLEKNL 157  
 DB 341 SFLEKDL 347  
 RESULT 14  
 AAY34897  
 ID AAY34897 standard; protein; 158 AA.  
 AC AAY34897;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 13-SEP-1999 (first entry)  
 XX Chlamydia pneumoniae transmembrane protein sequence.  
 DE  
 DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;  
 KW neutralising epitope.  
 XX Chlamydophila pneumoniae.  
 OS  
 OS WO9927105-A2.  
 PN  
 PN 03-JUN-1999.  
 PD  
 XX 20-NOV-1998; 98WO-IB001890.

XX 21-NOV-1997; 97FR-00014673.  
 PR 04-NOV-1998; 98US-0107078P.  
 XX (GEST) GENSET.  
 PA Griffais R;  
 PI  
 XX WPI; 1999-357842/30.  
 DR  
 XX Genome sequence of Chlamydia pneumoniae.  
 PT  
 XX Page 834; Disclosure; 1912pp; English.  
 PS  
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae. C.  
 CC pneumoniae causes respiratory disease such as pneumonia and bronchitis,  
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,  
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The  
 CC polypeptides encoded by the open reading frames of the C. pneumoniae  
 CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as  
 CC vaccines. Vectors containing C. pneumoniae nucleotides sequences can also  
 CC be used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17  
 CC -OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 158 AA;  
 Query Match 13.1%; Score 118; DB 2; Length 158;  
 Best Local Similarity 21.3%; Pred. No. 0.011;  
 Matches 32; Conservative 43; Mismatches 69; Indels 6; Gaps 3;  
 QY 27 IGVFNSQSIAMESEAAKAAQKLOS---EFGNEKTKLEKQAKDLQTKADDLQAKSAAMSN 83  
 DB 10 LGVNLKRCLESDLGKETEELKXQGFVKNAKIEBELTSYINKLQD-EDYMESLSD 68  
 QY 84 QAREDKQREPLELRNFEKSRDFAIRVEQAENTLRQVLAETIAAKKGLKLV 143  
 DB 69 SASEELKKKPFEDLSGEYNAVQSYQISQSNVNRKIQKLIQEVKIAAESVRSKEKLEAIL 128  
 QY 144 DSAGSGVMYLEKNLDTKEILEAINAAWK 173  
 DB 129 NEE--AVLAIPAGTDTKEIILNESPKK 156  
 RESULT 15  
 AAB21231  
 ID AAB21231 standard; protein; 717 AA.  
 AC AAB21231;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX Tomato LeMPF1.  
 DE  
 DE Tomato; MAR binding filament-like protein 1; MPF1; LeMPF1;  
 KW matrix attachment region; MAR; NtMPF1-2; anchor protein.  
 XX Lycopodium esculentum.  
 OS  
 OS WO200061615-A2.  
 PN  
 PN 19-OCT-2000.  
 PD  
 XX 12-APR-2000; 2000WO-US009723.  
 PF  
 XX 12-APR-1999; 99US-0128900P.  
 XX (DUPO) DU PONT DE NEMOURS & CO E I.  
 XX Harder PA, Meier I;  
 XX WPI; 2000-679464/66.

